LA-UR-16-21896
Approved for public release; distribution is unlimited.

Title: Whole genome landscapes of major melanoma subtypes

Author(s): Hayward, Nicholas; Wilmott, James; Waddell, Nicola; Johansson, Peter; Field, Matthew; Nones, Katia; Patch, AnnMarie; Kakavand, Hojabr; Alexandrov, Ludmil B.; Burke, Hazel; Jakrot, Valerie; Kazakoff, Stephen; Holmes, Oliver; Leonard, Conrad; Wood, Scott; Xu, Qinying; Waddell, Nick; Tembe, Varsha; Pupo, Gulietta; PaoliIseppi, Ricardo; Vilain, Ricardo; et al.

Intended for: Nature

Issued: 2016-03-21

# Whole genome landscapes of major melanoma subtypes

Nicholas K. Hayward[1,2]*, James S. Wilmott[1,3]*, Nicola Waddell[2,4]*, Peter A. Johansson[2]*, Matthew A. Field[5], Katia Nones[2,4], Ann-Marie Patch[2,4], Hojabr Kakavand[3], Ludmil B. Alexandrov[6], Hazel Burke[1], Valerie Jakrot[1], Stephen Kazakoff[2,4], Oliver Holmes[2,4], Conrad Leonard[2,4], Scott Wood[2,4], Qinying Xu[2,4], Nick Waddell[4], Varsha Tembe[7], Gulietta M. Pupo[7], Ricardo De Paoli-Iseppi[3], Ricardo E. Vilain[3], Ping Shang[3], Sarah-Jane Schramm[7], Antonia Pritchard[2], Ken Dutton-Regester[2], Anna Fitzgerald[8], Catherine A. Shang[8], Sean M. Grimmond[4,9], Jean Y. Yang[10], Jonathan R. Stretch[1], Richard F. Kefford[1,11], Peter Hersey[1,12], Georgina V. Long[1], Jonathan Cebon[13], Mark Shackleton[14], Andrew J. Spillane[1], Robyn P. M. Saw[1], John V. Pearson[2,4]*, John F. Thompson[1]*, Richard A. Scolyer[1,3,15]*, Graham J. Mann[1,7]*§

[1] Melanoma Institute Australia, University of Sydney, NSW, Australia

[2] QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia

[3] Discipline of Pathology, Sydney Medical School, The University of Sydney, Sydney, NSW, Australia

[4] Queensland Centre for Medical Genomics, Institute for Molecular Bioscience, University of Queensland, Brisbane, QLD, Australia

[5] Immunogenomics Laboratory, Australian National University, Canberra, ACT, Australia

[6] Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

[7] Centre for Cancer Research, Westmead Institute for Medical Research, University of Sydney, Westmead, NSW, Australia

[8] Bioplatforms Australia, Macquarie University, North Ryde, NSW, Australia

[9] Translational Research Centre, University of Glasgow, Glasgow, Scotland

[10] School of Mathematics and Statistics, University of Sydney, NSW, Australia

[11] Macquarie University, North Ryde, NSW, Australia

[12] Centenary Institute, University of Sydney, NSW, Australia

[13] Olivia Newton-John Cancer and Wellness Institute, Latrobe University, Austin Health, Heidelberg, Vic, Australia

[14] Peter MacCallum Cancer Centre and University of Melbourne, Melbourne, VIC, Australia

[15] Tissue Pathology and Diagnostic Oncology, Royal Prince Alfred Hospital, New South Wales, Australia

*The authors contributed equally

§ Correspondence should be addressed to: Graham J. Mann

**Summary**

**Cutaneous, acral and mucosal subtypes of melanoma were evaluated by whole-genome sequencing, revealing genes affected by novel recurrent mutations to the promoter (*TERT, DPH3*, *OXNAD1, RPL13A*, *RALY*, *RPL18A*, *AP2A1*), 5'-UTR (*HNRNPUL1*, *CCDC77*, *PES1*), and 3'-UTR (*DYNAP, CHIT1, FUT9, CCDC141, CDH9*, *PTPRT*) regions. *TERT* promoter mutations had the highest frequency of any mutation, but neither they nor *ATRX* mutations, associated with the alternative telomere lengthening mechanism, were correlated with greater telomere length. Genomic landscapes largely reflected ultraviolet radiation mutagenesis in cutaneous melanoma and provided novel insights into melanoma pathogenesis. In contrast, acral and mucosal melanomas exhibited predominantly structural changes, and mutation signatures of unknown aetiology not previously identified in melanoma. The majority of melanomas had potentially actionable mutations, most of which were in components of the mitogen-activated protein kinase and phosphoinositol kinase pathways.**

Melanoma is molecularly diverse and displays the highest mutation load of any cancer[1,2]. This high load has been considered attributable to ultraviolet radiation (UVR), with a C>T nucleotide transition signature dominating the remodelled somatic genome of the vast majority of cases[2]. However, some melanomas lack this signature and have relatively few point mutations[3,4]. Such potentially non-UVR driven melanomas are uncommon in cutaneous sites, but form the majority of those occurring in the eye, in mucosal surfaces, and on the hands and feet; the latter (acral) sites account for a much higher proportion of melanomas in Asians[5], than in Europeans. Elucidating the significant mutational processes and genomic drivers in all types of melanoma is important for its epidemiology and prevention and for identifying new therapeutic leads required for melanoma treatment globally.

The recent study of 333 melanomas by The Cancer Genome Atlas (TCGA) consortium[1] described molecular diversity of cutaneous melanoma and proposed a classification based on the commonest driver mutations: *BRAF*, *NRAS* and *NF1*. However, it had excluded acral, mucosal and eye melanomas. Further, the TCGA genomic analysis relied on exome and low-pass whole

genome sequencing, with thus a restricted examination of both the signatures of key mutational processes and of structural changes that may drive melanoma, and little assessment of driver mutations occurring outside the exome. Other sequencing studies have included some acral and mucosal melanomas[6,7], but have also been limited by small sample size and/or restriction to exome sequencing.

We therefore performed the first high coverage whole-genome sequencing study of a large cohort of melanomas, carefully selected to confirm their origin from cutaneous, acral or mucosal sites. The analysis revealed distinct mutation processes, profiles and types between acral and mucosal melanomas compared with cutaneous melanomas, a wide variety of non-coding mutations, and paradoxical relationships between telomere maintenance gene mutations and telomere length. Mutational signatures of UVR exposure were dominant in cutaneous melanomas, and complex structural rearrangements accounted for the majority of aberrations in acral and mucosal melanomas. This implies a pathogenesis for these tumours that is fundamentally different to most cutaneous melanomas, and suggests that therapies for them will need distinct targeting strategies.

**Study sample and approach**

Clinical, demographic and pathological details of the 183 patients and their tumours are presented in Supplementary Table 1. Samples comprised 75 primary melanomas, 93 melanoma metastases and 15 cell lines derived from melanoma metastases. Thirty-five melanomas were confirmed pathologically as acral (arising from the volar surface of the palms/soles of the hands/feet or subungual locations), eight were mucosal melanomas and the remainder were of cutaneous origin from non-glabrous skin, or had arisen from an unknown/occult primary. The latter were grouped with 'cutaneous' melanomas, since melanomas arising from unknown primary tumours have shown mutation profiles resembling cutaneous melanoma[8,9]. All samples underwent whole-genome sequencing to an average depth of 85x (range 43-219x) in the tumour sample and 64x (range 30-214x) in the matched normal (Supplementary Table 1)[55].

**Coding and non-coding mutation burden**

We identified a total of 20,894,255 substitutions (range 1,610 to 777,275 per tumour) and 96,467 small insertions/deletions (indels) (range 0 to 10,114 per tumour) at an average rate of 38.23 mutations per megabase (Mb) (range 0.54 to 260/Mb) (Supplementary Table 1 and Fig. 1). There were 156,770 coding mutations among them (range 19 to 5,536 per tumour), 0.8% of the total (Supplementary Table 2). Single nucleotide variant (SNV) and indel frequencies varied by melanoma subtype (t-test, $P < 1 \times 10^{-7}$), with >18 times more of these mutations in cutaneous melanomas (mean 49.17/Mb, range 0.71 to 259.69/Mb) than in acral and mucosal melanomas (mean 2.64/Mb, range 0.54 to 15.25/Mb) (Fig. 1b). Similarly, structural variants (SV) were more frequent in acral and mucosal melanomas (mean 342.40, range 6 to 1,148) than in cutaneous melanomas (mean 101, range 3 to 1,123) (t-test, $P < 1 \times 10^{-6}$) (Fig. 1c,d). Higher proportions of the acral and mucosal melanoma genomes were affected by copy number variation (CNV), and more regions were amplified (Mann-Whitney test, $P < 1.46 \times 10^{-4}$) than in cutaneous melanomas (Extended Data 1), and fewer of the cutaneous melanomas were aneuploid (chi-square test, $P = 0.049$).

**UVR and non-UVR mutagenesis mutation spectra**

Mutation spectra varied considerably by melanoma subtype (Fig. 2). The vast majority of cutaneous melanomas had a high proportion of C>T transitions (mean 81.3%; range 6.5 to 91.3%), most at dipyrimidines (mean 82.5%; range 21.3 to 92.4%), and CC>TT dinucleotide mutations (mean 1.4%; range 0.02 to 9.7%)(Fig. 2a). A mutational signature algorithm based on the sequence context of each mutation[2] identified 12 distinct mutational signatures (Fig. 2b, c). The genomes of most melanomas (total $n = 139$; 136 cutaneous) were dominated by three novel signatures of ultraviolet radiation (UVR) (signatures 7, 35, and 33, designated UV_i, UV_ii, UV_iii, respectively). These recapitulate the UVR mutational signature previously deduced from exome sequencing data[2] but, being derived from whole genome data, can now be resolved to disentangle the diverse DNA damaging properties of UV light. Briefly, the predominantly C>T substitutions at TpC dinucleotides (mutated base underlined) of signature UV_i is, based on context[10], most likely due to repair of 6,4-photoproducts. In contrast, signature UV_ii involves C>T substitutions at CpC and CpC which, from the sequence context, is characteristic of cyclobutane pyrimidine dimers[10].

Finally, signature UV_iii has high levels of T>C and T>A mutations that result from transitions and transversions, potentially due to indirect DNA damage after UV. Forty-four tumours, comprising 40 acral or mucosal melanomas and four cutaneous, all from less frequently sun-exposed sites, showed strikingly different mutation spectra from the UVR-dominant group (Fig. 2). Nine signatures were observed in this group[11]; six of them have been described in other cancer types, but never previously in melanoma and are presumably not the result of UVR exposure. The most frequent were signature 1, the result of an endogenous process initiated by spontaneous deamination of 5-methylcytosine, and signature 5 which is found ubiquitously and with unknown aetiology in cancers of many types[12] (Extended Data 2). Signature 13, attributed to activity of the AID/APOBEC family of cytidine deaminases was evident in 52 samples (28%) and was found in all melanoma subtypes. Signatures 8, 17 and 18, for which the aetiology is unknown, were also present in all melanoma subtypes at a combined frequency of 7-28%. Signature 6, associated with mismatch repair gene defects, was seen in one cutaneous melanoma, which had the highest value determined by MSIsensor[13].

**Overview of structural rearrangements**

The combined frequency of SV of all types, such as deletion, duplication, tandem duplication, and foldback inversion, was higher in acral and mucosal melanomas than in cutaneous melanomas ($P < 1 \times 10^{-6}$) (Fig. 1c and Supplementary Table 3). Acral and mucosal melanomas exhibited substantially more clusters of breakpoints, indicating more frequent complex rearrangements such as breakage-fusion-bridge and chromothripsis (Extended Data 3a). We detected complex rearrangements including breakage-fusion-bridge-like in 18% ($n = 32$) of tumours, most of which ($n = 20$, 63%) were acral or mucosal melanomas (Extended Data 4, 5).

**Overview of telomere length**

Whole-genome sequence data provided the first opportunity to assess the consequences of these small and large-scale mutagenic processes, and the driver mutation profiles of melanoma, on immortalisation via telomere length[14,15], which was estimated from using qmotif (see Methods). Telomere length varied widely among the 183 samples, ranging from 6-fold shorter to 10-fold

longer than the matched normal for each patient (Extended Data 6a). Telomere length was not correlated with melanoma subtype (Extended Data 6b), chromothripsis, or breakage-fusion-bridge events (Extended Data 3).

**Mutations in the non-coding genome**

All mutations were annotated for genomic context including regulatory, untranslated and protein coding regions, the first such comprehensive survey in melanoma, and one of the first in any cancer. *TERT* promoter mutations [16,17] were by far the most common, with 115/167 (69%) samples, and 86% of cutaneous melanomas mutated (Sanger sequencing) at one or more of four positions upstream of the initiation codon (Extended Data 6c). Mutations in each position created a new binding site for the ETS family transcription factor GABP[18], and they were mutually exclusive, with the exception of two samples with extremely high mutation burdens. These mutations were much less prevalent in acral and mucosal melanomas (4/38: 11%), which accounted for their association with lower SV count and higher SNV mutation burden (Fisher's exact test, $P < 1 \times 10^{-17}$). There was no association between *TERT* promoter and *BRAF* mutation, in contrast to previous reports[19] (chi-square test, $P > 0.09$).

Surprisingly, these putatively activating *TERT* promoter mutations were associated with modestly reduced telomere length (Mann-Whitney, $P = 0.0022$)(Extended Data 6d-f). Telomere elongation in cancer depends either on *TERT* or a process of alternative lengthening of telomeres (ALT)[20], often driven by mutations in *ATRX*[21]. The latter occur in a mutually exclusive pattern with *TERT* promoter mutations in glioblastoma[20]. Ten melanomas (all cutaneous) had predicted loss-of-function mutations in *ATRX,* including two nonsense mutations, but paradoxically nine of these tumours also carried mutations in the *TERT* promoter and there was no association between *ATRX* mutation and telomere length (data not shown). We therefore show that the hypothesis of immortalisation via mutational activation of *TERT* and facilitated telomere length extension is too simplistic, and that complex relationships likely exist between alternative telomere maintenance mechanisms, telomere length and immortalisation in melanoma.

Promoter mutations were also found frequently in *RPL13A*, *RALY*, *RPL18A*, *DPH3* and *OXNAD1*, each of which had multiple hotspot mutations (Supplementary Table 4). The designated promoter regions of *DPH3* and *OXNAD1* overlap and thus mutations in this region could affect expression of one or both genes. We found 39 tumours (21%) with *DPH3/OXNAD1* promoter mutations, all of which were cutaneous melanomas (Fisher's exact test, $P < 9 \times 10^{-5}$). The recurrently mutated positions (chr3: 16306504, chr3: 16306505 and chr3: 16306508; Extended Data 7a) are those reported previously to result in loss of ETS transcription factor binding sites[19] and, as in that study, *DPH3/OXNAD1* promoter mutations occurred preferentially (14/39; 36%) in melanomas with *NF1* mutations (Fisher's exact test, $P < 0.0003$).

Recurrent mutations in the promoter of *NFKBIE* have been reported in desmoplastic melanoma[22], and we identified such mutations in 10/183 (5%) tumours overall, including 3/9 desmoplastic melanomas. The most prevalent hotspot mutation (chr6: 44233400) occurred in five tumours, all cutaneous melanomas, but they lacked a desmoplastic melanoma component (Extended Data 7b). Mutations at this and one other position (chr6: 44233417) have been reported as mutated previously, but the other four sites are novel (Extended Data 7b) and presumably lay outside the captured exonic regions in the previous study. *NFKBIE* promoter mutations are thus enriched in desmoplastic melanomas (Fisher's exact test, $P < 0.0084$), but are potential drivers of a broader subset of cutaneous melanomas.

The most frequent 5'-UTR hotspot mutation, detected in 19/183 samples, was at the transcription start site of *RPS27* (chr1: 153963239 C>T)[23,24], in addition a dinucleotide mutation was detected in one patient (chr1: 153963239-153963240 CT>TC). *RPS27* also had several recurrent mutations in its promoter (Extended Data 7c), although none were predicted to result in gain or loss of transcription factor binding sites. Recurrent mutations also occurred frequently ($n > 11$) in the 5'-UTR of *MRPS31*, *RPS14*, *CHCHD2* and *PES1* (Supplementary Table 5). The latter was also identified as a common hotspot mutation by the TCGA Network[23]. That study did not report any 3'-UTR mutations; however, we observed recurrent 3'-UTR mutations frequently ($n > 5$) in *DYNAP, CHIT1, FUT9, CCDC141, CDH9* and *PTPRT* (Supplementary Table 6). None of these genes has

been associated with melanoma except *PTPRT*, which was recently reported to have frequent coding mutations[25].

**Significantly mutated genes**

The IntOGen[26] algorithm identified 127 ($q < 0.01$) significantly mutated genes (SMG), of which 26 (20%) were predicted to be driver genes, including previously reported melanoma oncogenes and tumour suppressor genes: *BRAF, NRAS, HRAS*, *TP53, CDKN2A, PTEN, NF1, DDX3X*, *RB1, ARID2*[3,23,27-31]. Mutation frequencies of key genes varied considerably between melanoma subtypes. For example, no acral or mucosal melanomas had mutations in *TP53, PTEN, DDX3X, RASA2, PPP6C, RAC1* or *RB1,* indicating that the molecular pathways driving these melanoma subtypes differ markedly from those driving cutaneous melanoma. Similarly, in line with previous studies[32], *KIT* mutations occurred more frequently in acral (3/35) and mucosal melanomas (2/8) than in cutaneous melanomas (6/140) (chi-square test, $P < 0.044$). The MutSig[33] algorithm ($q < 0.01$) identified only three SMG (*BRAF*, *NRAS*, *TP53*). A comparison of the SMG detected here with those reported by TCGA[23] is presented in Supplementary Table 7. The distribution of mutations in some key melanoma driver genes is shown in Fig. 3a. SV and CNV contributed heavily to the overall mutation frequency of some melanoma driver genes, especially *NF1*, *TP53*, *PTEN* and *KIT* (Extended Data 1b,c; Fig. 3b). Notably, the majority of samples with *BRAF* hotspot mutations had amplifications of the gene, which also tended to co-occur with amplifications of *MET*, located close to it on chromosome 7. Similarly, *KIT* and *PDGFRA,* co-located within 0.35 Mb on chromosome 4, were frequently co-amplified. *MITF* amplifications were present in 10% of samples and were distributed across all three melanoma subtypes. Many acral and mucosal melanomas had SV and high-level amplifications on the long arm of chromosome 11, often targeting *CCND1* (Extended Data 3c).

**TCGA genomic subgroups of melanoma**

We analysed the characteristics of the genomic subgroups of cutaneous melanoma proposed by the TCGA study: *BRAF* mutated, *RAS* mutated, *NF1* mutated or triple wild-type.[23]

*BRAF mutated samples*

*BRAF* somatic non-synonymous mutations were found in 85 of the 183 samples (46%) (Extended Data 8). V600E substitutions accounted for 48 (56%) of these mutations. Other activating mutations at codons 600 and 601 were less frequent (V600K, *n* = 17; V600R, *n* = 2*;* V600D, *n* = 1; K601E, *n* = 2). These hotspot mutations were mutually exclusive with those in *NRAS* (test for independence $P < 0.0001$, Fisher's exact test) whereas other *BRAF* mutations were sometimes found in conjunction with *RAS* (*N/H/K*) or *NF1* mutations (Fig. 3a). Significantly fewer *BRAF* hotspot mutations were observed in acral (*n* = 6, 17%) and mucosal melanomas (*n* = 0) than in cutaneous melanomas (*n* = 64, 46%) (chi-square test, $P < 0.005$).

*RAS mutated samples*

*NRAS* was the most commonly mutated member of the RAS gene family, with 51 (28%) of samples carrying mutations, 49 of which occurred at 'hotspot' codons 12, 13 and 61 (Extended Data 9). *HRAS* (*n* = 8) and *KRAS* (n = 4) mutations were less frequent, and notably, only five of the *HRAS* mutations and none of the *KRAS* mutations occurred at the commonly activating hotspot codons of these genes. More *NRAS* hotspot mutations were found in cutaneous (*n* = 43, 31%) and mucosal (*n* = 2, 25%) melanomas than in acral melanomas (*n* = 4, 11%) (Fisher's exact test, $P < 0.013$).

*NF1 mutated samples*

*NF1* aberrations occurred in 32/183 (17%) samples; 23 had point mutations/small indels only (15 nonsense, four missense, three splice site and one frame-shift) (Extended Data 8). Of the 32 samples with *NF1* mutations, 14 (44%) had potentially biallelic aberrations. The high contribution of SVs to mutation load of *NF1* observed in the current study indicates that previous studies in which SVs were not taken into account under-estimated the mutation burden in *NF1* and consequently over-estimated the proportion of triple wild-type melanomas. *NF1* mutated melanomas had the highest mutation load (mean of 76.27/Mb), more than double that of *NF1* wild-type samples (mean of 30.17/Mb) (t-test, $P = 0.0009$). *NF1* mutations were inversely correlated with hotspot *BRAF* mutations (Fisher's exact test, $P = 1.4 \times 10^{-5}$) but not hotspot *RAS* mutations. *NF1* was mutated in

22 (68%) of non-hotspot *BRAF/NRAS* melanomas (14 cutaneous and 8 acral samples). Moreover *NF1* aberrations were found to co-occur with *RASA2* mutations (Fisher's exact test, $P < 0.003$), as recently described[29,34].

*Triple wild-type melanomas*

Most mucosal and acral ($n$ = 22/43, 51%) melanomas, but only 15/140 (11%) cutaneous melanomas were triple wild-type. However, mutations in other cancer driver genes were frequent in this group, including loss-of-function mutations in *CDKN2A, TP53* and *ARID2*, and activating hotspot mutations in: *GNAQ*, *SF3B1,* and *CTNNB1.* In particular, *GNAQ* and *SF3B1* mutations, commonly mutated in uveal melanoma[35,36] but rarely in cutaneous melanoma[23], were found in four of the six triple wild-type mucosal melanomas: one *GNAQ* and three *SF3B1* hotspot mutations. These were the only samples in the entire cohort that carried activating hotspot mutations in these genes, implying a hitherto unappreciated common set of driver genes and pathway aberrations underlying development of mucosal and uveal melanomas. Seven triple wild-type melanomas carried mutations in *KIT*, five of which also had high-level amplifications. Triple wild-type melanomas had significantly more focal amplifications of *KIT*, *CCND1*, *MDM2* and *KRAS* than non-triple wild-type melanomas (Fisher's exact test, $P < 5.4 \times 10^{-5}$, $1.8 \times 10^{-5}$, 0.018, 0.003, respectively).

**Gene fusions**

Gene fusions targeted several genes previously associated with melanoma development, specifically: *RAF1*[37], *RAC1*[3]; *MAP2K2*[38]. Others included: *GRM3*[39], *MAP3K9*[30], *TRRAP*[31] and *PLCB4*[31] (Supplementary Table 3). Fusions also targeted known melanoma tumour suppressor genes, namely *NF1* ($n$ = 3), *PARK2* ($n$ = 2), *TP53*, *PTEN*, and *PPP6C*. A *RAF1* gene fusion has previously been shown to cause increase phosphorylation of endogenous MEK, when compared with cells expressing wild-type RAF1[40]. We observed *RAF1* gene fusions in two cutaneous samples, one of which triple wild-type and the other *NF1*-mutated. The same *RAF1* exons were retained in both fusions but they were fused to different genes (*CDH3* and *GOLGA4*) (Extended Data 9).

Interestingly, fusions involved several kinase genes (Supplementary Table 3), the most frequently affected being *PAK1* (*n* = 5 samples), which has not previously been identified as frequently mutated in melanoma. Four of the five observed *PAK1* fusions occurred in acral melanomas, two of which were triple wild-type. The one cutaneous melanoma *PAK1* fusion carried a BRAF V600E mutation. These *PAK1* fusions are unlikely to be activating events, as the kinase domain was retained in only four of five samples, in each case there was a different gene fusion partner (*AQP11*, *FCHSD2*, *TENM4*, *OR9Q1*) and the samples each contain additional breakpoints (two to four breaks per sample) within the *PAK1* gene. The next most frequently fused kinase gene was *DGKB*, with two fusions in acral triple wild-type melanomas and one in a cutaneous BRAF V600E mutated sample. In addition to the fusion events each sample harboured multiple breakpoints within the *DGKB* gene, again suggesting that the fusions are unlikely to result in a gain of function. This notion is supported by the finding of 20 SNV in *DGKB* across 14 samples, with five of the samples potentially having bi-allelic mutations (Supplementary Table 2) and three of this group having clear loss-of-function mutations (two nonsense and one splice mutation). Taken together, these data point to *DGKB* being a key novel tumour suppressor gene in melanoma.

**Common pathway alterations**

To identify recurrently targeted molecular pathways in melanoma, we analysed protein-affecting mutations in all tumours and annotated them by KEGG pathway to identify those with significantly enriched mutation load (Supplementary Table 8). The MAPK signalling pathway was enriched 1.34-fold (Fisher's exact test, $P < 1 \times 10^{-51}$) and was altered in 179 of 183 (98%) samples. Representative frequencies of the main components of the MAPK pathway mutated in this cohort are shown in Fig. 4, which also highlights other frequently mutated signalling pathways in melanoma. To compare pathways driving melanomas of different subtypes, we assessed gene and pathway aberrations in Fig. 4a according to cutaneous or non-cutaneous (acral and mucosal) melanomas. Both groups were dominated by mutations to the MAPK pathway and RTKs. Non-cutaneous melanomas harboured fewer alterations in the other pathways, and most aberrations were SV (33/38, 87%).

## Clinical implications

The frequency of clinically actionable mutations was assessed by annotating genomic variants using the Intogen Cancer Drivers Actionability (ICDA) database (2014), which identifies mutations that may confer sensitivity to therapeutic agents[41]. A high proportion of cutaneous (138/140, 99%) and non-cutaneous (39/43, 91%) melanomas harboured mutations in at least one gene that potentially conferred sensitivity to an agent that is FDA-approved or currently in a clinical trial in any cancer (Extended data 10). With the inclusion of CNV all but one cutaneous melanoma carried at least one such aberration, and even this case contained *PPP2R5C* and *CTNNB1* mutations, which are potentially targetable[41].

The aberration frequencies of some potentially actionable genes were significantly higher in cutaneous melanomas than in the non-cutaneous melanomas, including *BRAF* (54% and 21%, respectively), *CDKN2A* (32% and 5%), *TP53* (23% and 0%), *NRAS* (32% and 14%), *HDAC9* (25% and 7%) and *MTOR* (9% and 0%) (two-proportion z-test, all $P < 0.05$). Conversely, lower rates of mutations occurred in cutaneous than non-cutaneous melanomas for *NF1* (17% and 26%, respectively), *KIT* (9% and 16%), *CCND1* (0% and 7%), *STK4* (2% and 7%) and *MDM2* (2% and 7%); however, only *CCND1* reached statistical significance (two-proportion z-test, $P = 0.002$).

As observed in previous studies, age at primary melanoma diagnosis was correlated with TCGA genomic subgroup (*BRAF*/*RAS*/*NF1*/triple wild-type): *BRAF* mutation was associated with the youngest age at diagnosis (median 53 years, Tukey honest significant difference test, all $P < 0.03$), *NF1* mutation with the oldest age at diagnosis (median 80 years), while *NRAS* mutant (median 59 years) and triple wild-type (median 65 years) were intermediate. However, only a weak correlation was observed in cutaneous melanomas between age at diagnosis and total mutation load ($R = 0.28$, $P = 0.0015$). Shorter survival since primary melanoma diagnosis was associated with *BRAF* mutation (Log-rank test, $P = 0.017$), but not with overall load of SNV or SV. *TERT* mutations were almost equally prevalent in primary (65%) and metastatic (72%) melanomas, whether or not non-cutaneous samples were excluded (87% and 85%, respectively), suggesting these mutations arise

early in tumour development. *TERT* mutation status was not associated with overall survival, whether analysed by stage at diagnosis, or restricted to cutaneous melanoma. Power was limited to detect associations with *TERT* mutation, because of their high prevalence, in contrast with previous studies in less heavily UVR exposed cohorts[42].

## Discussion

This is the largest analysis of melanoma using high-coverage whole-genome sequencing to date and the largest to compare cutaneous, acral and mucosal subtypes at the genomic level. Acral and mucosal melanomas showed a dramatically different genomic landscape from cutaneous melanoma, with a far lower mutation burden, dominated by large-scale structural changes. We observed novel mutagen signatures in cutaneous melanomas that likely represent as yet uncharacterised mechanisms of UVR damage. The principal mutation mechanisms driving mucosal and most acral melanomas were, in contrast, not attributable to UVR and may result from novel carcinogenic exposures that are in common with cancers other than melanoma and non-melanoma skin cancer. Notably, some mucosal melanomas exhibited driver mutations (*GNAQ*, *SF3B1*) previously considered characteristic of uveal melanoma. In further contrast to cutaneous melanomas, acral and mucosal melanomas were not frequently driven by *TP53*, *PTEN* or *RB1* pathway lesions, and while a high proportion showed predominant MAPK signalling pathway mutations, they exhibited a variety of “triple wild type” mechanisms such as *KIT* mutations and focal amplifications of *KIT*, *CCND1*, *MDM2* and *KRAS*. Novel recurrently fused genes were also identified (*PAK1*, *DGKB*). Taken together, a high proportion of both cutaneous and non-cutaneous melanomas exhibited mutations in genes for which targeted therapies exist or are being evaluated in clinical trials.

A large number of genes exhibited non-coding mutations, surveyed here for the first time on any scale in melanoma. Considering only those affecting the 5’-UTR, we observed recurrent mutations in *RPS14*, *CHCHD2* and *PES1* (5’-UTR), in addition to the genes we and the TCGA Network have previously reported as mutated (*RPS27*, *MRPS31*); and we report for the first time genes affected by recurrent 3’-UTR mutations, namely *DYNAP, CHIT1, FUT9, CCDC141, CDH9* and *PTPRT.* The latter mutations were within microRNA binding sites in only three instances, none of which affected

the critical seed region, so this mechanism is unlikely to be functionally relevant to them. We observed novel promoter mutations in *DPH3*, *OXNAD1, RPL13A*, *RPL18A* and *RALY,* and extended the spectrum and associations of *NFKBIE* promoter mutation beyond desmoplastic melanoma.

Finally, we observed a striking lack of correlation between either *TERT* promoter mutation or *ATRX* mutation and greater telomere length, which calls into question the simplest hypothesis about the relationship of these mutations to tumorigenesis, namely that they facilitate telomere length maintenance, whether via telomerase or the ALT mechanism. The high prevalence of *TERT* promoter mutations, in particular, must be accounted for by selective advantages they confer on melanoma via the dysregulation of telomere protection, not simply via telomere length.

## Methods

### Human melanoma samples

The fresh-frozen tissue and blood samples analysed in the current study were obtained from Australian melanoma biospecimen banks, including the Melanoma Institute Australia (MIA) (n=160), QIMR-Berghofer Institute of Medical Research (QIMR) (n=15), Ludwig Institute for Cancer Research (n=4), Peter MacCallum Cancer Centre/Victorian (n=4) biospecimen banks. All tissues and bloods form part of prospective collection of fresh-frozen samples accrued with written informed patient consent and institutional review board approval. Fresh surgical specimens were macro-dissected and tumour tissues were procured (with as little contaminating normal tissue as possible) and snap frozen in liquid nitrogen within 1 hour of surgery. All samples were pathologically assessed prior to inclusion into the study, with samples requiring greater than 80% tumour content and less than 30% necrosis to be included. All samples were independently reviewed by expert melanoma pathologists RAS and REV to confirm the presence of melanoma and qualification of the above criteria. Samples requiring tumour enrichment underwent macrodissection or frozen tissue coring (Cryoxtract, Woburn MA USA) using a marked H&E slide as a reference.

The histopathology of all mucosal and acral samples was reviewed by RAS to confirm the diagnosis. Acral melanomas were classified as occurring within acral skin of the palm of the hand, sole of the foot and under nail beds. The lack of hair follicles, thickened stratum corneum and clinical site was confirmed in all cases. Mucosal melanomas were defined as occurring in the mucosal membranes lining oral, respiratory, gastrointestinal and urogenital tracts. The H&E slides of the primary melanomas were reviewed for all mucosal and acral samples and any tumour that arose in the junction of the acral/mucosal and cutaneous skin was excluded.

### DNA extractions

Tumour DNA was extracted using DNeasy Blood and Tissue Kits (69506, Qiagen Ltd), according to the manufacturer's instructions. Blood DNA was extracted from whole blood using Flexigene DNA Kits (51206, Qiagen Ltd). All samples were quantified using the NanoDrop (ND1000,

Thermoscientific) and Qubit® dsDNA HS Assay (Q32851, Lifetechnologies) and DNA size and quality were tested using gel electrophoresis. Samples with a concentration of less than 50 ng/µl, or absence of a high molecular weight band in electrophoresis gels, were excluded from further analyses.

**Whole genome sequencing**

Whole-genome sequencing was performed on Illumina Hiseq 2000 instruments (Illumina, San Diego, CA, USA) at three Australian sequencing facilities (Australian Genomic Research Facility, Ramaciotti Centre for Genomics, John Curtin School of Medical Research) and Macrogen (Geumcheon-gu, Seoul, South Korea). All facilities performed library construction using TruSeq DNA Sample Preparation kits (Illumina) as per Illumina instructions. The subsequent 100 bp pair-end libraries were sequenced using Truseq SBS V3-HS kits.

**Whole genome sequence processing and quality control**

Sequence data was aligned to the GRCh37 assembly using multi-threaded BWA 0.6.2-mt resulting in sorted lane level files in sequence alignment/mapping (SAM) format which were compressed and converted to binary file (BAM) created by samtools 0.1.19. Sample-level merged BAMs, one each for matched germline and tumour samples were produced by in-house tools and duplicate reads marked with Picard MarkDuplicates 1.97 (http://picard.sourceforge.net). Quality assessment and coverage estimation was carried out by qProfiler and qCoverage (http://sourceforge.net/projects/adamajava). To test for the presence of sample or data swaps all sequence data were assessed for concordance at approximately 1.4 million polymorphic genomic positions including the genotyping array data by qSignature. All BAM files are available in the EGA (Accession number: EGAS00001001552).

**Telomere length estimation**

To count motifs in genome sequencing files, such as telomere repeats, we used the qmotif tool. For increased speed, qmotif uses a 2-stage matching system where the first stage is a quick-but-strict string match and the second stage is a slower but more flexible regular expression match.

Only reads that pass stage 1 go on to the much slower match in stage 2. For the purpose of telomere quantification, we typically use a string that represents 3 concurrent repeats of the canonical telomere motif (TTAGGGTTAGGGTTAGGG) as the stage 1 match and a simple pattern match for stage 2 which captures any read with 2 or more concurrent occurrences of the telomeric repeat with variation allowed in the first 3 bases. qmotif is implemented in JAVA using the Picard library (version 1.110). qmotif is driven by a single plain-text configuration file in the "Windows INI-file" style. The input is a whole genome sequencing BAM file that has been duplicate-marked and coordinate-sorted. qmotif is multi-threaded, requiring 8 cores and 20 GB of memory to run efficiently. More details on the configuration file and the required inputs can be found on the qmotif wiki page: https://sourceforge.net/p/adamajava/wiki/qMotif.

**Somatic Mutation Analysis**

Somatic mutations and germline variants were detected using a dual calling strategy using qSNP[43] and GATK[44] and indels of 1–50 bp in length were called with Pindel[45] and GATK. All mutations were submitted to the International Cancer Genome Consortium (ICGC)[46] Data Coordination Centre (DCC). Mutations were annotated with gene consequence using Ensembl gene annotation with SnpEff. Somatic genes which were significantly mutated were identified with IntOGen[26].

Somatic copy number and ploidy were determined using the TITAN tool[47]. Somatic structural variants were identified using the qSV tool and chromosomes containing highly significant non-random distributions of breakpoints were identified as previously described[48,49]. Chromosomes identified to have clustering of breakpoints were inspected against criteria for chromothripsis[50] and the breakage-fusion-bridge cumulative rearrangement model[51]. Chromosomes with high numbers of translocations were identified with a minimum threshold of 10 translocation breakpoints per chromosome following manual review.

Mutational signatures were predicted in each sample using a published framework[2,52]. Essentially the substitution mutations across the whole genome in all cases were analysed in context of the

flanking nucleotides (96 possible trinucleotide combinations). Identified signatures were compared to other validated signatures and the frequency of each signature per Mb was determined.

**Body site distribution of SNV/indel and SV**

Total numbers of SNV/Indel and SV were compared according to primary melanoma body site, categorized into abdomen, acral hand, acral foot, back, lower arm, lower leg, mucosal, neck, shoulder, thorax, upper arm, upper leg and face and scalp. Any samples with an unknown primary site or occult classification were excluded from analysis. Heat maps were produced in Spotfire-Tibco (version 6.0, spotfire.tibco.com) based on the combined total number of SNV and indels, or by SV. A two-colour heat map (red high, blue low) was produced with the average number of mutations set to white. Colours were overlaid using Illustrator CS6 (Adobe) onto an adapted SVG human body diagram obtained from Commons Wikimedia (https://commons.wikimedia.org/wiki/File:Human_body_front_and_side.svg).

**Clinically actionable genes**

We aligned mutations with the Intogen Cancer Drivers Actionability (ICDA) database (2014)[41] to identify genes that confer sensitivity to therapeutic agents. Firstly, the Intogen OncodriverRole was used to assign either activating or a loss of function role to mutated genes using Spotfire-Tibco[53]. Loss of heterozygosity, silent mutations, deletions to activating genes or amplifications to loss of function genes were not included in the analysis. Remaining variants were then assigned the ICDA tiered “Protein Druggability” classifications that group therapeutic agents based on availability of an FDA-approved or clinical trial agent in use in any cancer. Any variant that was directly or indirectly targeted with an FDA-approved or clinical trial therapeutic was used in further analysis. The proportion of tumours with a mutation to a particular actionable gene was calculated and classified based on mutation type into: 1) SNV/indel, 2) SNV/indel and SV, 3) SV or 4) copy number variation only.

**Commonly mutated genes and pathways**

We hand-curated a list of commonly mutated tumour suppressor genes and oncogenes and analysed their mutation frequencies. Mutations were defined as SNV/indel (missense mutation, nonsense mutation, RNA, splice site, frame-shift insertion, frame-shift deletion, in-frame deletion, translation start site and non-stop mutation), structural variants, copy number amplifications and copy number deletions (as defined by GISTIC). We overlaid non-silent mutations and indels in protein-coding regions onto pathways defined by KEGG and GO gene sets from MSigDB v5.0. Genes shown in Fig. 4a were assigned signaling pathways and then ranked based on the proportion of samples with a mutation. A pathway was considered altered in a given sample, if at least one gene in the pathway contained an SNV/indel or SV. The pathways were then stratified according to cutaneous or mucosal and acral subtypes. A mutation file with samples IDs and their mutated pathways was entered for analysis into the OncoPrinter tool in the cBio Cancer Genomics Portal (http://cbioportal.org)[54] .

## References


1 Akbani, R. *et al.* Genomic Classification of Cutaneous Melanoma. *Cell* **161**, 1681-1696 (2015).

2 Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415-421 (2013).

3 Krauthammer, M. *et al.* Exome sequencing identifies recurrent somatic RAC1 mutations in melanoma. *Nature genetics* **44**, 1006-1014 (2012).

4 Berger, M. F. *et al.* Melanoma genome sequencing reveals frequent PREX2 mutations. *Nature* **485**, 502-506 (2012).

5 Chi, Z. *et al.* Clinical presentation, histology, and prognoses of malignant melanoma in ethnic Chinese: a study of 522 consecutive cases. *BMC cancer* **11**, 85 (2011).

6 Furney, S. J. *et al.* Genome sequencing of mucosal melanomas reveals that they are driven by distinct mechanisms from cutaneous melanoma. *The Journal of pathology* **230**, 261-269 (2013).

7 Furney, S. J. *et al.* The mutational burden of acral melanoma revealed by whole-genome sequencing and comparative analysis. *Pigment cell & melanoma research* **27**, 835-838 (2014).

8 Dutton-Regester, K. *et al.* Melanomas of unknown primary have a mutation profile consistent with cutaneous sun-exposed melanoma. *Pigment cell & melanoma research* **26**, 852-860 (2013).

9 Egberts, F. *et al.* Metastatic melanoma of unknown primary resembles the genotype of cutaneous melanomas. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO* **25**, 246-250 (2014).

10 Ravanat, J. L., Douki, T. & Cadet, J. Direct and indirect effects of UV radiation on DNA and its components. *Journal of photochemistry and photobiology. B, Biology* **63**, 88-102 (2001).

11 Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering signatures of mutational processes operative in human cancer. *Cell reports* **3**, 246-259 (2013).

12 Alexandrov, L. B. & Stratton, M. R. Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. *Current opinion in genetics & development* **24**, 52-60 (2014).

13 Niu, B. *et al.* MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. *Bioinformatics* **30**, 1015-1016 (2014).

14 Mardin, B. R. *et al.* A cell-based model system links chromothripsis with hyperploidy. *Molecular systems biology* **11**, 828 (2015).

15 Nones, K. *et al.* Genomic catastrophes frequently arise in esophageal adenocarcinoma and drive tumorigenesis. *Nature communications* **5**, 5224 (2014).

16 Horn, S. *et al.* TERT promoter mutations in familial and sporadic melanoma. *Science* **339**, 959-961 (2013).

17 Huang, F. W. *et al.* Highly recurrent TERT promoter mutations in human melanoma. *Science* **339**, 957-959 (2013).

18 Bell, R. J. *et al.* Cancer. The transcription factor GABP selectively binds and activates the mutant TERT promoter in cancer. *Science* **348**, 1036-1039 (2015).

19 Fredriksson, N. J., Ny, L., Nilsson, J. A. & Larsson, E. Systematic analysis of noncoding somatic mutations and gene expression alterations across 14 tumor types. *Nature genetics* **46**, 1258-1263 (2014).

20 Bryan, T. M., Englezou, A., Dalla-Pozza, L., Dunham, M. A. & Reddel, R. R. Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines. *Nature medicine* **3**, 1271-1274 (1997).

21 Ramamoorthy, M. & Smith, S. Loss of ATRX Suppresses Resolution of Telomere Cohesion to Control Recombination in ALT Cancer Cells. *Cancer cell* **28**, 357-369 (2015).

22 Shain, A. H. *et al.* Exome sequencing of desmoplastic melanoma identifies recurrent NFKBIE promoter mutations and diverse activating mutations in the MAPK pathway. *Nature genetics* **47**, 1194-1199 (2015).

23 Cancer Genome Atlas, N. Genomic Classification of Cutaneous Melanoma. *Cell* **161**, 1681-1696 (2015).

24 Dutton-Regester, K. *et al.* A highly recurrent RPS27 5'UTR mutation in melanoma. *Oncotarget* **5**, 2912-2917 (2014).

25 Ding, L. *et al.* Clonal architectures and driver mutations in metastatic melanomas. *PloS one* **9**, e111153 (2014).

26 Gonzalez-Perez, A. *et al.* IntOGen-mutations identifies cancer drivers across tumor types. *Nature methods* **10**, 1081-1082 (2013).

27 Andersen, L. B. *et al.* Mutations in the neurofibromatosis 1 gene in sporadic malignant melanoma cell lines. *Nature genetics* **3**, 118-121 (1993).

28 Hodis, E. *et al.* A landscape of driver mutations in melanoma. *Cell* **150**, 251-263 (2012).

29 Krauthammer, M. *et al.* Exome sequencing identifies recurrent mutations in NF1 and RASopathy genes in sun-exposed melanomas. *Nature genetics* **47**, 996-1002 (2015).

30 Stark, M. S. *et al.* Frequent somatic mutations in MAP3K5 and MAP3K9 in metastatic melanoma identified by exome sequencing. *Nature genetics* **44**, 165-169 (2012).

31 Wei, X. *et al.* Exome sequencing identifies GRIN2A as frequently mutated in melanoma. *Nature genetics* **43**, 442-446 (2011).

32 Curtin, J. A., Busam, K., Pinkel, D. & Bastian, B. C. Somatic activation of KIT in distinct subtypes of melanoma. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **24**, 4340-4346 (2006).

33 Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214-218 (2013).

34 Arafeh, R. *et al.* Recurrent inactivating RASA2 mutations in melanoma. *Nature genetics* (2015).

35 Harbour, J. W. *et al.* Recurrent mutations at codon 625 of the splicing factor SF3B1 in uveal melanoma. *Nature genetics* **45**, 133-135 (2013).

36 Van Raamsdonk, C. D. *et al.* Frequent somatic mutations of GNAQ in uveal melanoma and blue naevi. *Nature* **457**, 599-602 (2009).

37 Palanisamy, N. *et al.* Rearrangements of the RAF kinase pathway in prostate cancer, gastric cancer and melanoma. *Nature medicine* **16**, 793-798 (2010).

38 Nikolaev, S. I. *et al.* Exome sequencing identifies recurrent somatic MAP2K1 and MAP2K2 mutations in melanoma. *Nature genetics* **44**, 133-139 (2012).

39 Prickett, T. D. *et al.* Exon capture analysis of G protein-coupled receptors identifies activating mutations in GRM3 in melanoma. *Nature genetics* **43**, 1119-1126 (2011).

40 Jones, D. T. *et al.* Oncogenic RAF1 rearrangement and a novel BRAF mutation as alternatives to KIAA1549:BRAF fusion in activating the MAPK pathway in pilocytic astrocytoma. *Oncogene* **28**, 2119-2123 (2009).

41 Rubio-Perez, C. *et al.* In Silico Prescription of Anticancer Drugs to Cohorts of 28 Tumor Types Reveals Targeting Opportunities. *Cancer Cell* **27**, 382-396 (2015).

42 Griewank, K. G. *et al.* TERT Promoter Mutation Status as an Independent Prognostic Factor in Cutaneous Melanoma. *JNCI Journal of the National Cancer Institute* **106**, dju246 (2014).

43 Kassahn, K. S. *et al.* Somatic point mutation calling in low cellularity tumors. *PloS one* **8**, e74380 (2013).

44 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* **20**, 1297-1303 (2010).

45 Ye, K., Schulz, M. H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**, 2865-2871 (2009).

46 Hudson, T. J. *et al.* International network of cancer genome projects. *Nature* **464**, 993-998 (2010).

47 Ha, G. *et al.* TITAN: inference of copy number architectures in clonal cell populations from tumor whole-genome sequence data. *Genome research* **24**, 1881-1893 (2014).

48 Patch, A. M. *et al.* Whole-genome characterization of chemoresistant ovarian cancer. *Nature* **521**, 489-494 (2015).

49 Waddell, N. *et al.* Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* **518**, 495-501 (2015).

50 Korbel, J. O. & Campbell, P. J. Criteria for inference of chromothripsis in cancer genomes. *Cell* **152**, 1226-1236 (2013).

51 Kinsella, M. & Bafna, V. Combinatorics of the breakage-fusion-bridge mechanism. *Journal of computational biology : a journal of computational molecular cell biology* **19**, 662-678 (2012).

52 Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415-421 (2013).

53 Schroeder, M. P., Rubio-Perez, C., Tamborero, D., Gonzalez-Perez, A. & Lopez-Bigas, N. OncodriveROLE classifies cancer driver genes in loss of function and activating mode of action. *Bioinformatics* **30**, i549-i555 (2014).

54 Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* **6**, pl1 (2013).

55 Wilmott, J.S. et al. Tumour procurement, DNA extraction, coverage analysis and optimisation of mutation-detection algorithms for human melanoma genomes. *Pathology*. 2015 Oct 29. [Epub ahead of print]

## Supplementary Information

Supplementary Table 1 | Clinical and mutation Data

Supplementary Table 2 | Coding mutations (SNV and indel) in melanoma (MAF)

Supplementary Table 3 | Structural rearrangements in melanoma

Supplementary Table 4 | Gene promoters frequently mutated in melanoma

Supplementary Table 5 | Recurrent 5' UTR mutations in melanoma

Supplementary Table 6 | Recurrent 3' UTR mutations in melanoma

Supplementary Table 7 | Significantly mutated genes

Supplementary Table 8 | Perturbed pathways in melanoma

## Acknowledgments

This work was supported by Melanoma Institute Australia (MIA), Bioplatforms Australia, New South Wales Ministry of Health, Cancer Council New South Wales, Program Grants of the National Health and Medical Research Council of Australia (NHMRC) and Cancer Institute New South Wales, and the Australian Cancer Research Foundation. NKH, RAS and KD-R are supported by NHMRC Fellowships. KN is supported by a Keith Boden Fellowship. MS is supported by Pfizer Australia, the Victorian Endowment for Knowledge, Science and Innovation and NHMRC. Sample collection was supported by MIA, the Victorian Cancer Agency, the Victorian Cancer Biobank, Victorian State Government Operational Infrastructure Support Program, the Melanoma Research Alliance and the Melbourne Melanoma Project. LBA is supported by a J. Robert Oppenheimer Fellowship at Los Alamos National Laboratory. The research used resources provided by the Los Alamos National Laboratory Institutional Computing Program, which is supported by the US Department of Energy National Nuclear Security Administration under Contract No. DE-AC52-06NA25396. Research performed at Los Alamos National Laboratory was carried out under the auspices of the National Nuclear Security Administration of the United States Department of Energy. The authors gratefully acknowledge the support of colleagues at MIA, Royal Prince Alfred Hospital, NSW Health Pathology, the Westmead Institute for Medical Research, and of Dr Andreas Behren, Olivia Newton-John Cancer and Wellness Institute, and thank Doug Stetner for computing assistance.

## Author Contributions

PAJ, MAF, KN, AMP, LBA, AP, SK, OH, CL, SW, QX, NMW and KDR analysed genomic data; VJ, PS, HK, RdP, VT, GMP and HB collected, prepared and analysed samples and data; AF, CAS, JYY and SJS contributed to design, planning and direction; JVP, NW and SMG developed and directed the data analysis pipeline; JFT, MS, JC, JRS, RFK, PH, GVL, AJS, RPMS and REV collected samples and data; NKH, JSW, PAJ, NW, RAS and GJM designed and directed the study, analysed data and wrote the paper. All authors discussed the results and reviewed the manuscript.

## Author Statement

The authors declare no competing financial interests. BAM files are available in the EGA (Accession number: EGAS00001001552). Correspondence and requests for materials should be directed to GJM (graham.mann@sydney.edu.au).

## FIGURES AND LEGENDS

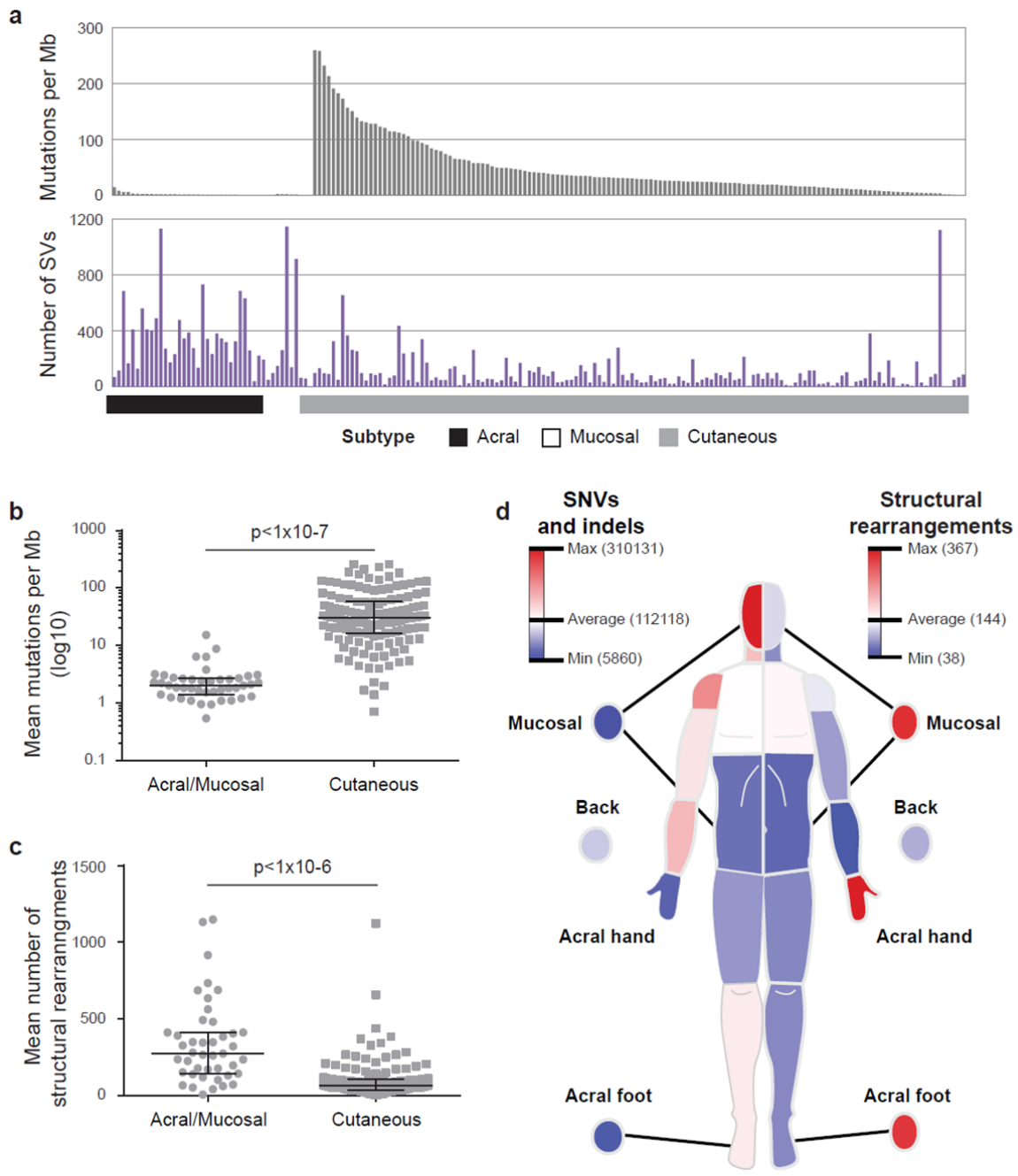



**Figure 1 | Mutation burden in melanoma. a,** The frequency of somatic substitution and indel mutations varied across the cohort, with a mean mutation rate across all tumours of 38.23 mutations/Mb, range 0.54 to 260/Mb. The mean frequency of structural rearrangements across all tumours was 158, range 3 to 1,148. **b,** There were significant differences in substitution and indel counts by melanoma subtype (t-test $P < 1 \times 10^{-7}$), with acral and mucosal melanomas (mean 2.64/Mb, range 0.54 to 15.25/Mb) carrying more than 18-fold fewer mutations than cutaneous

melanomas (mean 49.17/Mb, range 0.71 to 259.69/Mb). **c,** The acral and mucosal melanomas (mean 342.40, range 6 to 1,148) showed significantly (t-test $P < 1 \times 10^{-6}$) higher total numbers of structural variants compared with cutaneous melanomas (mean 101, range 3 to 1,123). **d,** Schematic representation of relative abundance of substitutions, indels and structural rearrangements by body site of the antecedent primary tumour. The highest (red) average substitution and indel counts occur in sun-exposed areas of the face scalp, neck and shoulders. The inverse is true for the structural variants depicted on the right side of the body, with the highest counts occurring in sun-shielded body sites of mucosal and acral origin.


**Figure 2 | Mutational processes in melanoma. a,** The proportion of somatic base changes is indicated for each tumour. The melanomas segregate into two groups with the vast majority of the cutaneous melanomas in the left group, which shows a high proportion of C>T transitions. Mutational signature analysis identified 12 distinct signatures. Each signature's contribution to a tumour's mutation burden is displayed as **b**, mutation per Mb and **c**, percentage. Samples in the left group, predominantly cutaneous melanomas, were dominated by ultraviolet radiation (UVR) signature. Three novel UVR signatures (termed UV_i, UV_ii, and UV_iii) were identified, which are a refinement of the UVR signature previously deduced from exome sequence data. Of the 139 samples with a UVR signature, UV_i was dominant in 134. One sample was dominated by signature U_ii and four samples formed a cluster of similar pattern with high levels of signature 2, UV_i, and UV_iii. Among the cutaneous melanomas, desmoplastic melanomas showed a higher proportion of signature UV_ii (Mann-Whitney test $P = 0.014$), possibly reflecting differences in UVR exposure and different types of DNA adducts formed. Signature 6, associated with mismatch repair, was identified in one sample (yellow). The tumours on the right, with low mutation burden, were dominated by signature 1 (associated with age) and signature 5 (aetiology unknown), found ubiquitously in cancers of diverse types, as well as signature 8, which has previously been observed in medulloblastoma and breast cancer.

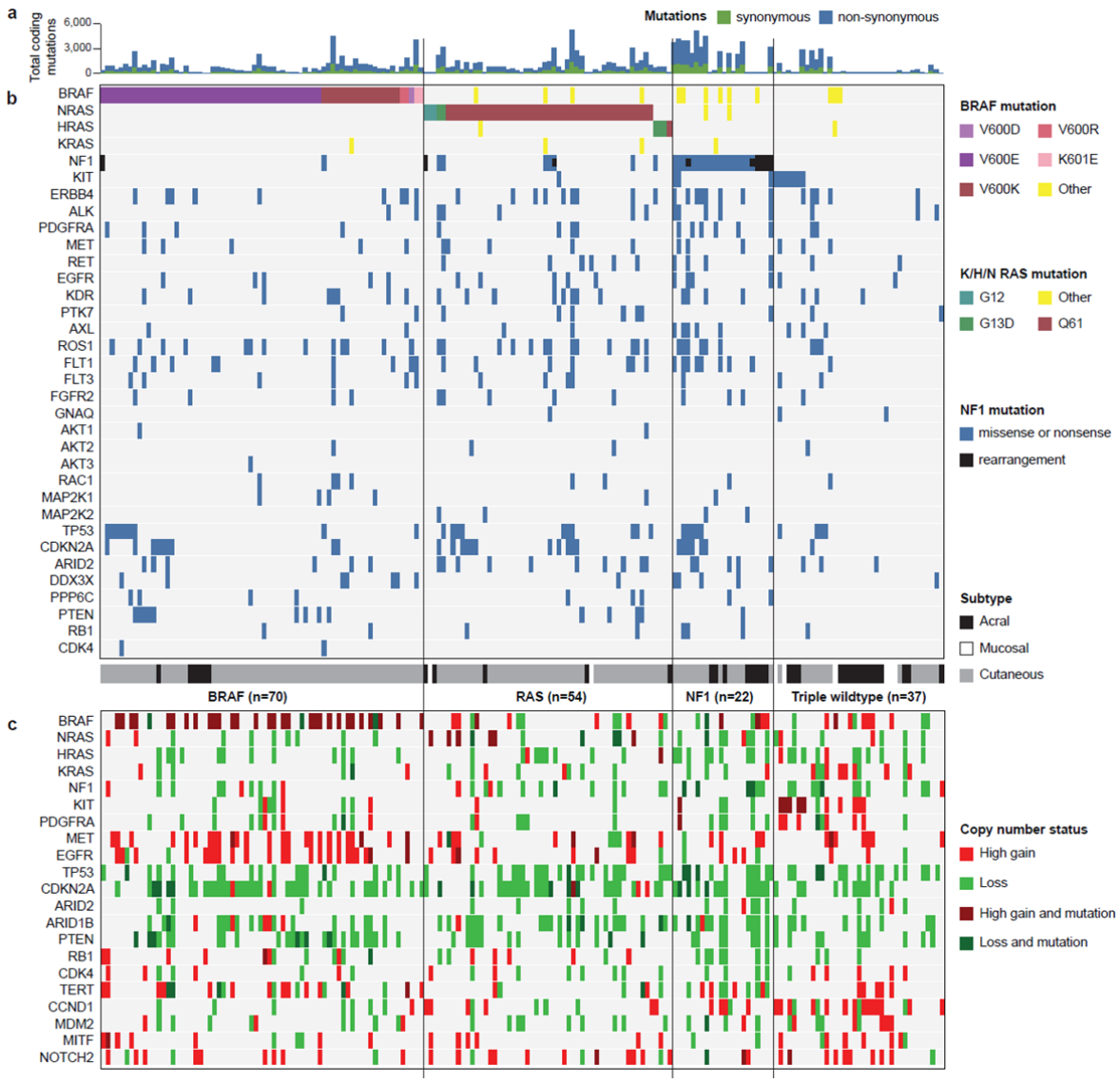



**Figure 3 | Mutations and copy number changes in driver genes. a,** Histogram shows the somatic coding mutation rate for the 183 samples. **b,** Mutations in driver genes with the most frequent oncogenic mutations in the *BRAF* and *RAS* genes coloured by mutation type. The melanoma subtype is shown at the bottom. **c,** copy number changes in melanoma associated genes are shown for homozygous loss, loss of 1 copy and high gain (≥6 copies). The melanoma subtype is shown at the bottom.

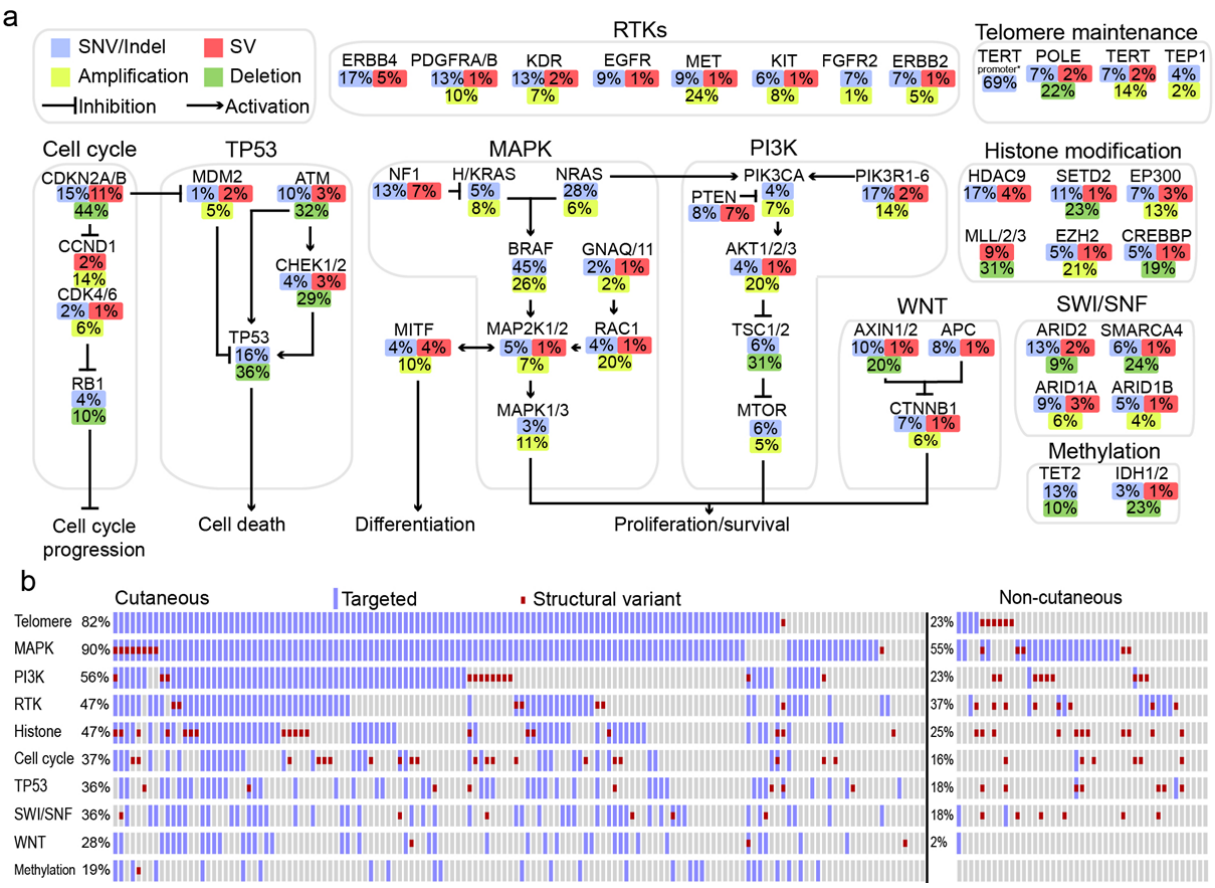
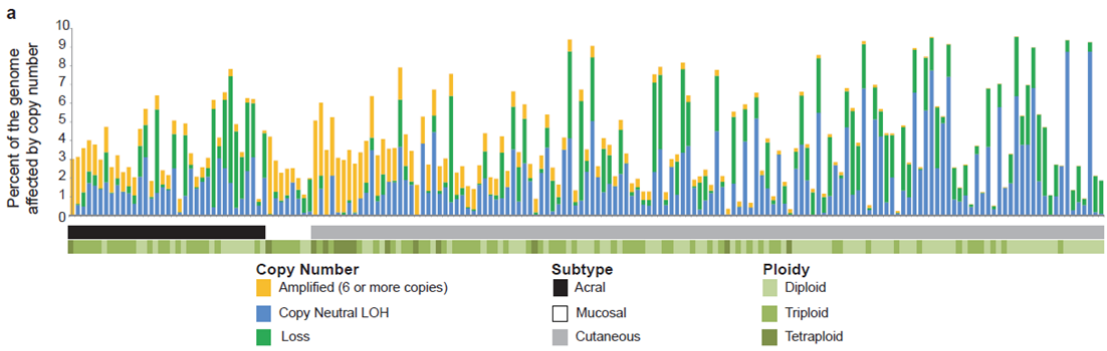
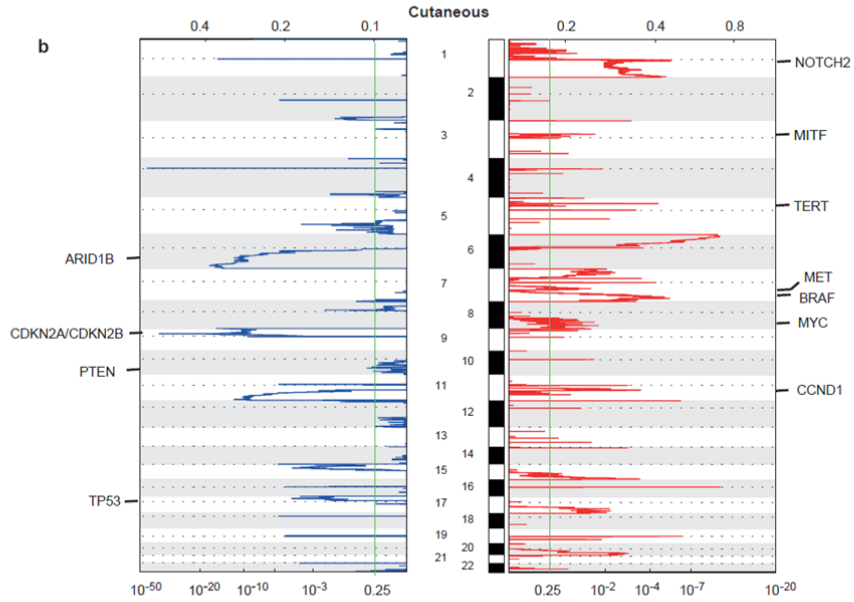
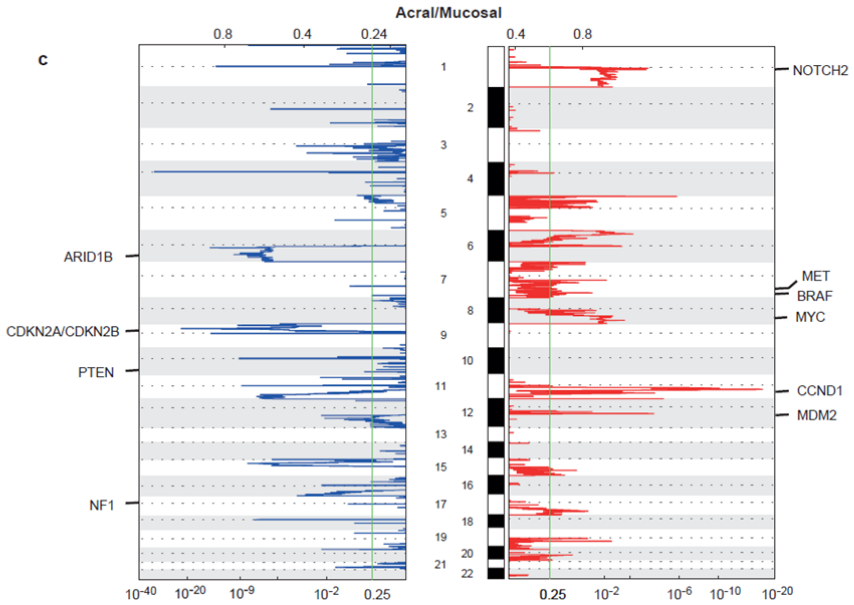



**Figure 4 | Genes and signaling pathways recurrently altered in melanoma. a,** Percentage of all patients with a non-silent mutation or indels in a protein coding region (blue), structural variant (red) and copy number variation (high-level amplification orange and homozygous deletion green) to a melanoma driver gene. **b,** Pathway OncoPrint of cutaneous and non-cutaneous melanomas derived from the manually curated genes and pathways from the figure above. Frequency of alterations to the respective pathways is expressed as a percentage of either cutaneous (n = 140) or non-cutaneous melanomas (n = 43). Pathways were classified as altered if a gene within the pathway contained a non-silent mutations/indels (blue) or structural variants (red).

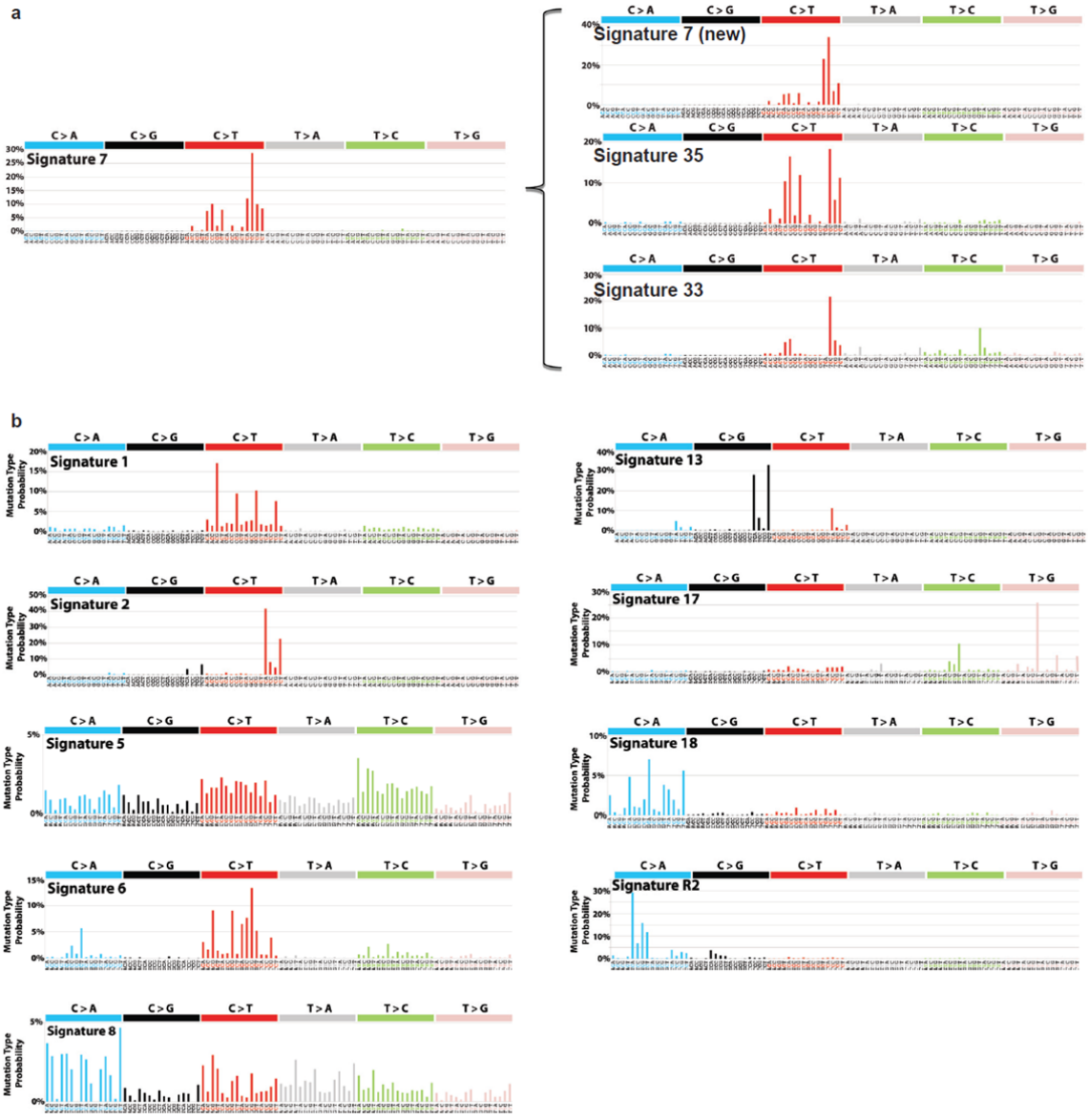
# EXTENDED DATA FIGURES AND LEGENDS



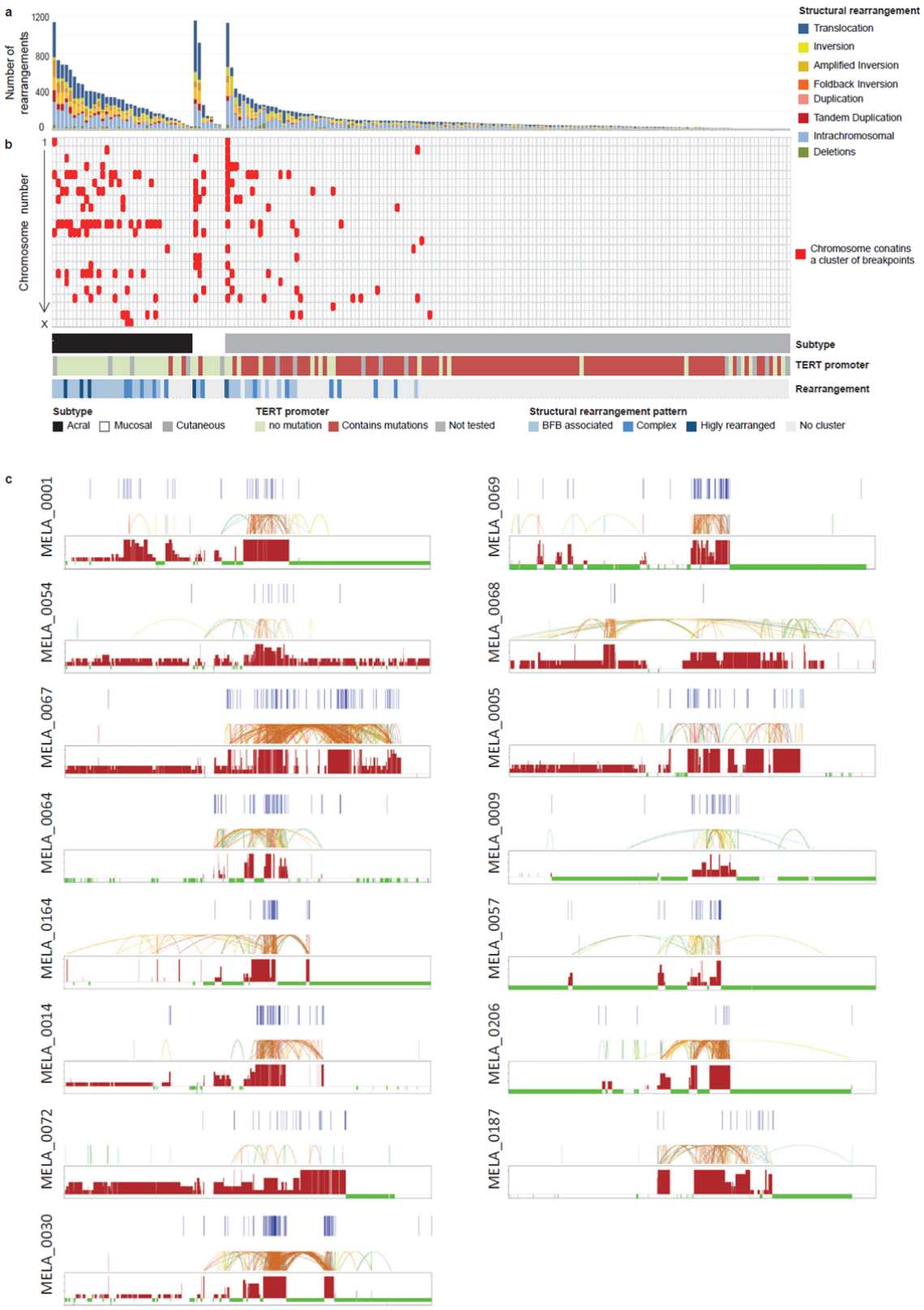



**Extended Data 1 | Copy number and ploidy in melanoma. a.** The proportion of each of the melanoma genomes that were affected by loss (copy number 1 or 0), copy neutral loss of heterozygosity and high gain (≥6 copies) are shown in the histogram. The melanoma subtype and degree of ploidy are illustrated in the colour bar beneath the histogram. GISTIC analysis was performed to determine significant regions of recurrent copy number change in the **b** cutaneous and **c** acral and mucosal tumours.



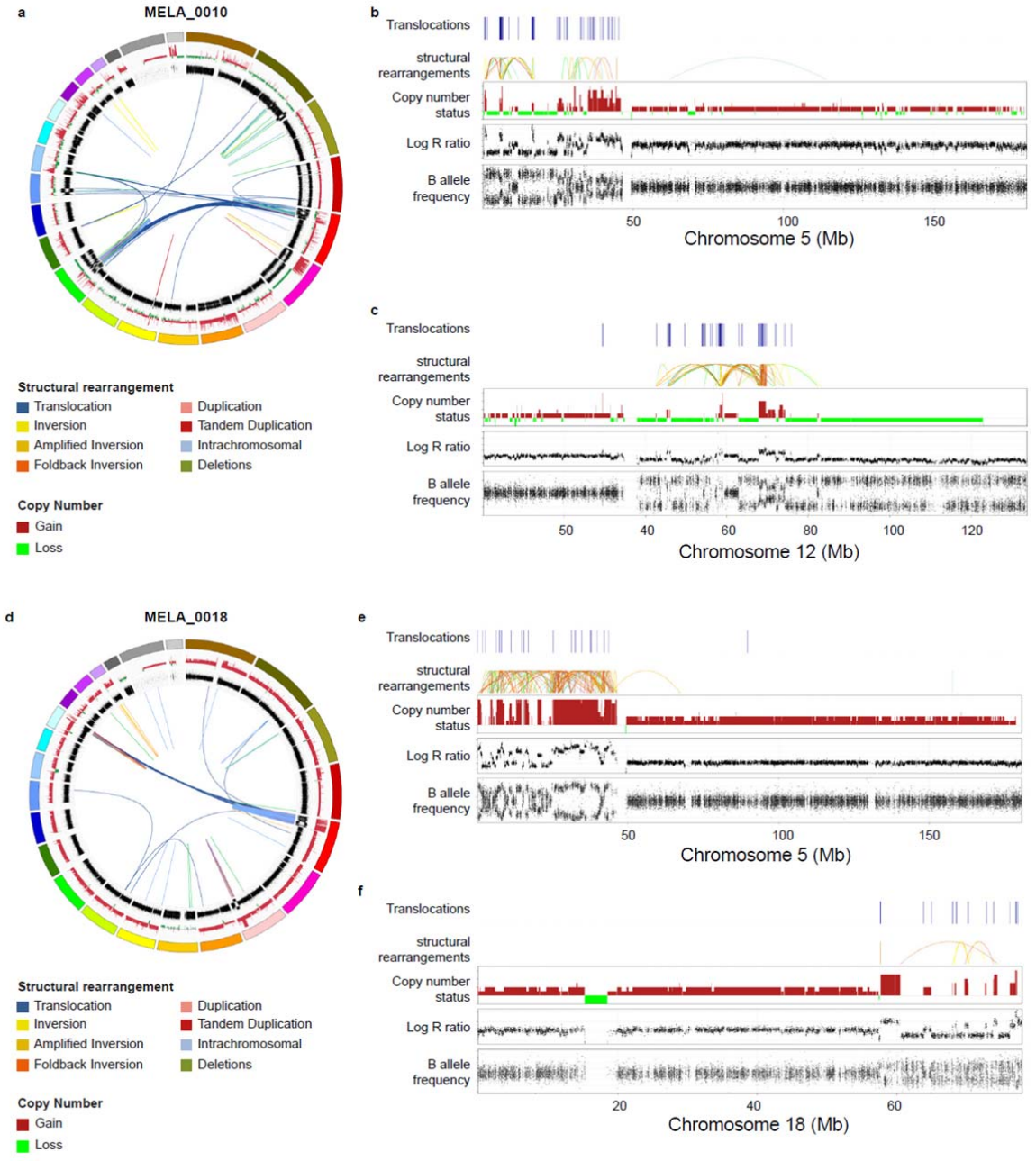

**Extended Data 2 | Mutation signatures in melanoma.** Twelve mutation signatures in melanoma were identified taking into account the sequence context immediately before and after the mutation. Each signature is displayed showing the probability for each of the 96 mutation types. **a,** Three novel signatures, associated with ultraviolet radiation (UVR), were identified (termed UV_i, UV_ii, and UV_iii). These signatures perfectly recapitulate the UVR mutation signature previously extracted from exome sequencing data. Signature UV_i is predominately C>T transitions occurring at TpC dinucleotides and, based on similarity with the sequence context, it is most likely due to the

formation of 6,4-photoproducts. Signature UV_ii is described by C>T transitions at CpC dinucleotides. This sequence context is characteristic for the formation of cyclobutane pyrimidine dimers due to UVR exposure. Signature UV_iii has, besides high proportion of C>T substitutions, high levels of T>C and T>A mutations demonstrating the ability of UVR to generate both transitions and transversions. The underlying processes damaging DNA and resulting in signature UV_iii are currently unknown but they could potentially be due to indirect DNA damage. **b,** Of the nine non-UVR associated signatures, three (signature 1, 5 and 17) have previously been observed in melanoma and six have been seen in other cancer types but not previously noted in melanoma. Signature 1 is the result of an endogenous mutational process initiated by spontaneous deamination of 5-methylcytosine and is associated with age of the patient. Signature 5 (aetiology unknown) was the dominant signature in tumours with no UVR signature; it is together with signature 1 the most common signature and been observed in a diverse set of cancer types. Signature 17 (aetiology unknown) has been seen in melanoma previously and was here observed in 52 tumours (exclusively cutaneous and acrals). Signature 8 was one of the dominant signatures in tumours with no UVR signature (together with signature 5 and 1); it has previously been observed in breast cancer and medulloblastoma and was evident in 39 melanomas. Signatures 2 and 13 have been attributed to been attributed to activity of the AID/APOBEC family of cytidine deaminases and are often observed in the same samples; here most non-UV samples showed both signatures, but highest level of signature 2 was observed in a cluster of fours samples in which signature 2, UV_i, and UV_iii were observed. Signature 6 is associated with defective DNA mismatch repair and typically found in microsatellite unstable tumours; it was observed in one melanoma in this cohort. Signature 18 has been observed frequently in neuroblastoma; it was identified in 13 melanomas. Signature R2 reflects a sequencing artefact.



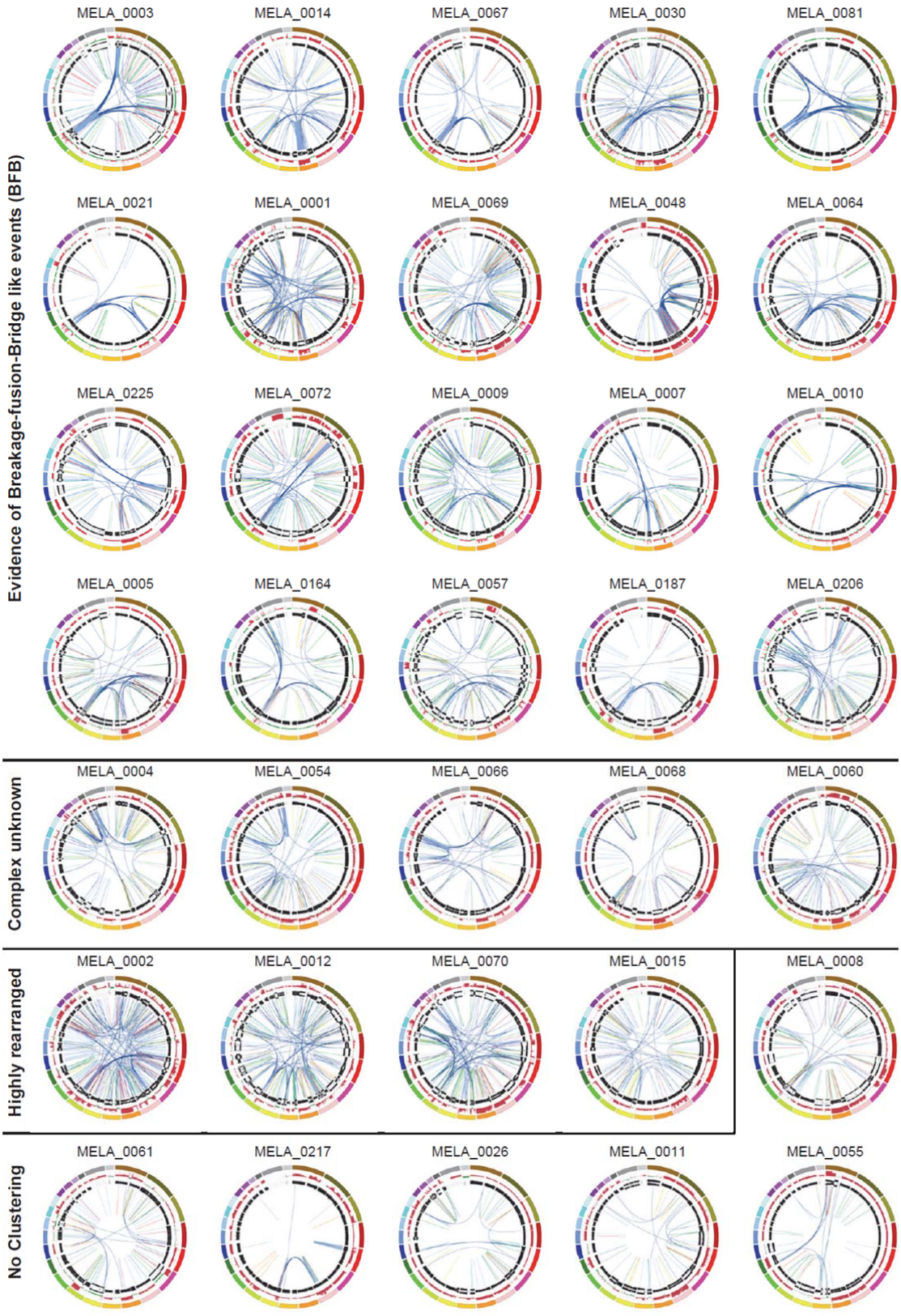

**Extended Data 3 | Structural rearrangements: overview and cluster on chromosome 11 in acral melanoma. a,** The number and type of structural rearrangements is shown for each tumour (n=183). **b,** Chromosomes containing complex rearrangements are shown in red defined as

chromosomes with a clustered arrangement of breakpoints (Kolmogorov-Smirnov goodness of fit test $P < 0.0001$) and a high density of breakpoints (number of breakpoints exceeds the extreme threshold value of the upper quartile plus 5* inter-quartile range for the chromosomal distribution of breakpoints). The observed pattern of breakpoint clusters and copy number events was used to classify tumours as breakage-fusion-bridge associated, complex, highly rearranged or no clusters of events. *TERT* promoter mutations are also shown. **c,** Patterns of structural rearrangements and copy number in all acral melanomas which contained a cluster of structural rearrangements on chromosome 11. Graphs show from the most frequent translocations, other structural rearrangements, copy number (red=gain, green=loss).

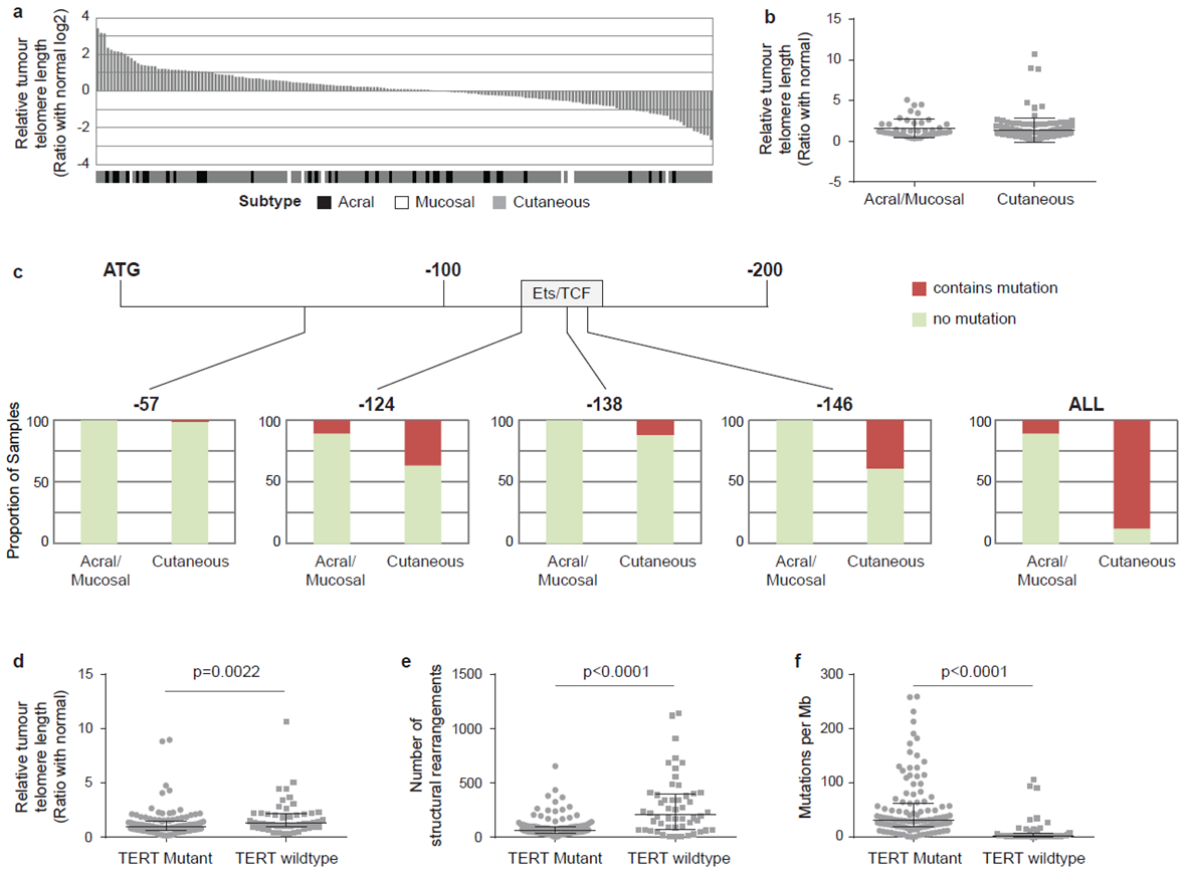



**Extended Data 4 | Fine structure of two rearrangements. a,** Circos plot of MELA_0010 shows a cluster of structural rearrangements on chromosomes 5 and 12. The outer coloured ring on the Circos plot represents each chromosome, the next ring is the copy number (red = gain and green = loss), the next ring is the B-allele frequency, and the structural rearrangements are shown by the lines in the centre of the plot. **b,** The cluster of rearrangements on chromosome 5 in MELA_0010

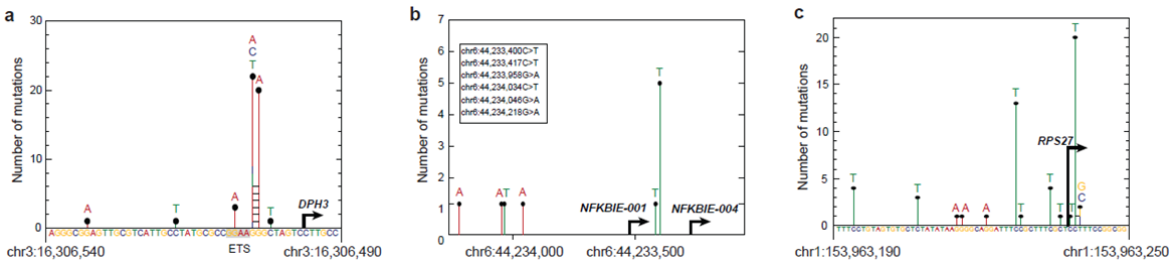
are shown from top to bottom: translocations (blue), all other rearrangements, copy number status, logR ratio and B allele frequency. **c,** The cluster of rearrangements on chromosome 12 in MELA_0010 shows evidence of breakage-fusion-bridge (BFB) with loss of telomeric region and inversions with increased copy number. **d.** Circos plot of MELA_0018 shows a cluster of structural rearrangements on chromosomes 5 and 18. The cluster of rearrangements in MELA_0018 on chromosome 5 (**e**) and chromosome 18 (**f**) are shown from top to bottom: translocations, all other rearrangements, copy number status, logR ratio and B allele frequency.


Evidence of Breakage-fusion-Bridge like events (BFB)
MELA_0003
MELA_0014
MELA_0067
MELA_0030
MELA_0081
MELA_0021
MELA_0001
MELA_0069
MELA_0048
MELA_0064
MELA_0225
MELA_0072
MELA_0009
MELA_0007
MELA_0010
MELA_0005
MELA_0164
MELA_0057
MELA_0187
MELA_0206
Complex unknown
MELA_0004
MELA_0054
MELA_0066
MELA_0068
MELA_0060
Highly rearranged
MELA_0002
MELA_0012
MELA_0070
MELA_0015
MELA_0008
No Clustering
MELA_0061
MELA_0217
MELA_0026
MELA_0011
MELA_0055

**Extended Data 5 | Complex structural rearrangements are frequent in acral melanomas.** Circos plots are shown for all 35 acral melanomas in the study. The outer coloured ring on the Circos plot represents each chromosome, the next ring is the copy number (red = gain and green = loss), the next ring is the B-allele frequency, and the structural rearrangements are shown by the lines in the centre of the plot. Tumours that contain breakage-fusion-bridge events, complex rearrangements, highly rearranged genomes or those that contain no clusters of structural rearrangements are indicated.

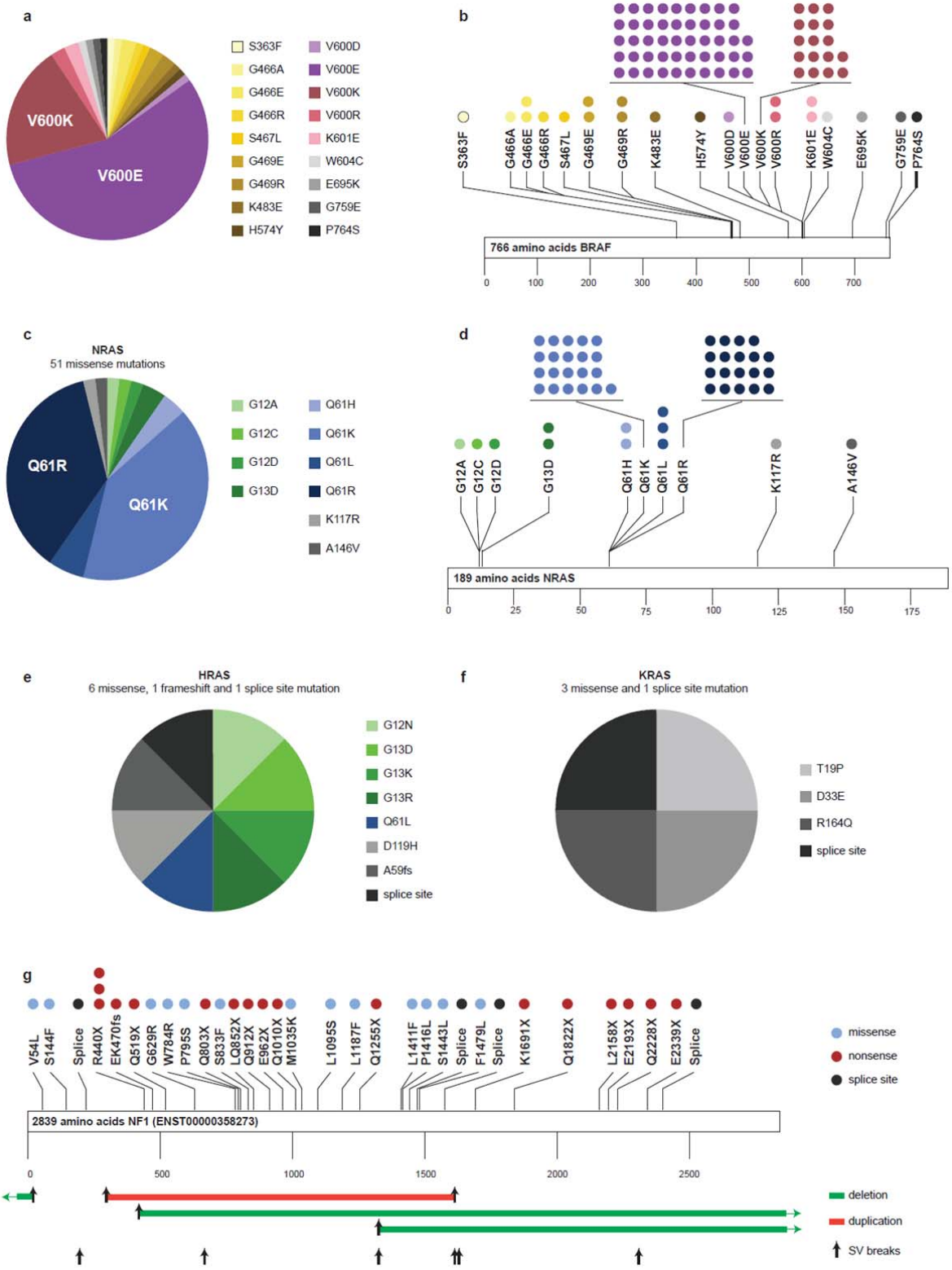



**Extended Data 6 | Telomere length and *TERT* promoter mutations. a,** The relative melanoma telomere length in each tumour compared to the matched normal. **b,** The telomere length was not associated with melanoma subtype. **c,** Somatic mutations in the promoter of the *TERT* gene were detected by whole-genome sequencing and/or capillary sequencing. Mutations were found at -57, -124, -138 and -146 from the ATG site. Tumours with TERT promoter mutations were associated with (Mann-Whitney test) **d,** telomere length; **e,** the number of structural rearrangements; and **f,** the number of mutations per Mb.




**Extended Data 7 | Recurrent *DPH3*, *NFKBIE* and *RPS27* promoter mutations.** Height and colour of sticks indicate recurrence and mutant base, respectively. Tandem mutations are indicated with black horizontal bars. **a,** Recurrent mutations in promoter of *DPH3* at loci 8 (n=20), 9 (n=22), and 12 (n=3) basepairs upstream from the transcription start site (TSS). The ETS transcription factor family core motif is shaded in gray. **b,** Recurrent mutation (n=5) at chr6:44,233,400 in first exon of *NFKBIE-001* isoform and 129 bases upstream from TSS of *NFKBIE-004*, which typically is the expressed transcript in most tissue including melanoma. Four novel mutations were identified further 500kb upstream. **c,** Recurrent mutations on and upstream from the TSS of *RPS27*.

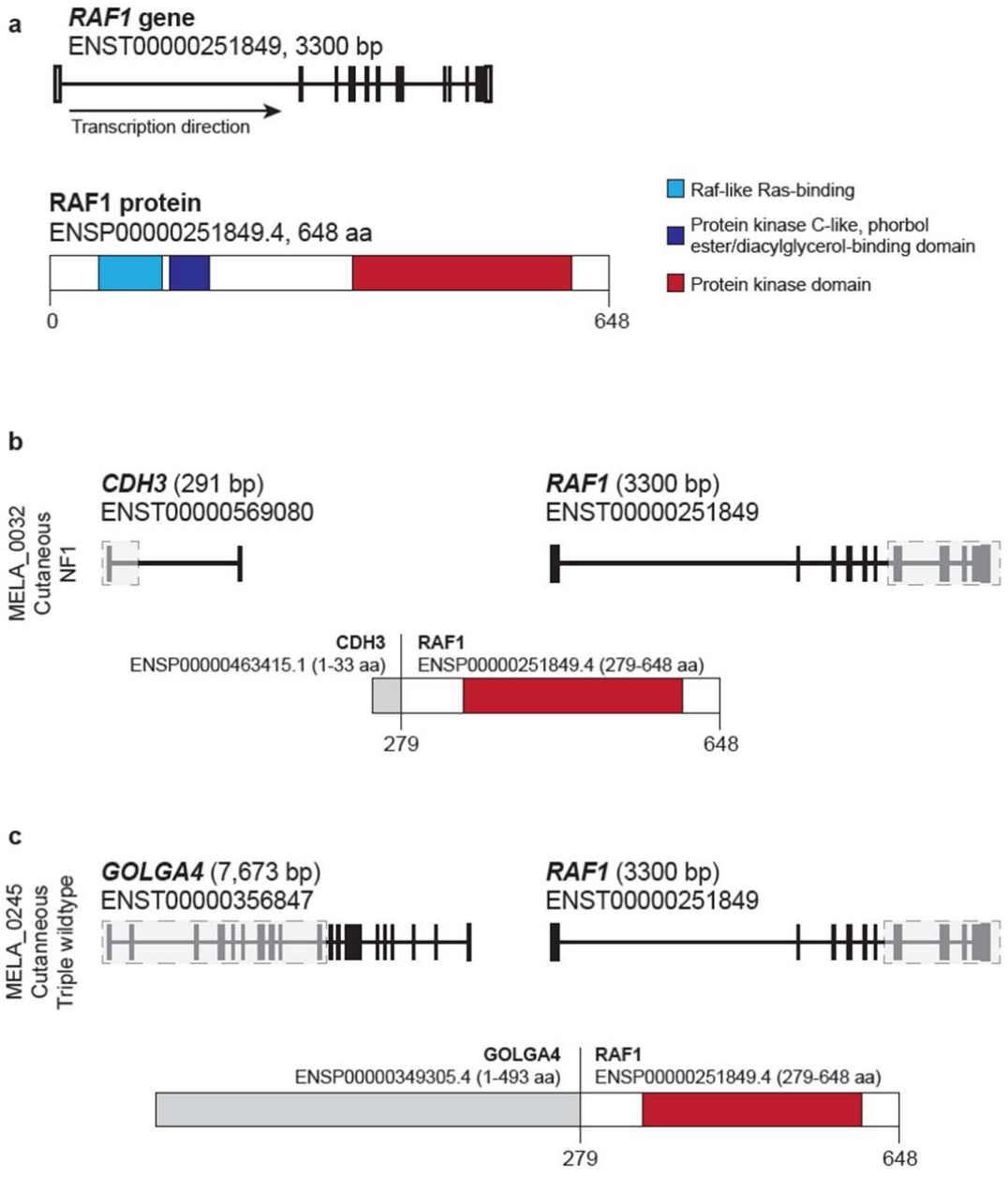
1


a
V600K
V600E
S363F
G466A
G466E
G466R
S467L
G469E
G469R
K483E
H574Y
V600D
V600E
V600K
V600R
K601E
W604C
E695K
G759E
P764S
b
S363F
G466A
G466E
G466R
S467L
G469E
G469R
K483E
H574Y
V600D
V600E
V600K
V600R
K601E
W604C
E695K
G759E
P764S
766 amino acids BRAF
0
100
200
300
400
500
600
700
c
NRAS
51 missense mutations
Q61R
Q61K
G12A
G12C
G12D
G13D
Q61H
Q61K
Q61L
Q61R
K117R
A146V
d
G12A
G12C
G12D
G13D
Q61H
Q61K
Q61L
Q61R
K117R
A146V
189 amino acids NRAS
0
25
50
75
100
125
150
175
e
HRAS
6 missense, 1 frameshift and 1 splice site mutation
G12N
G13D
G13K
G13R
Q61L
D119H
A59fs
splice site
f
KRAS
3 missense and 1 splice site mutation
T19P
D33E
R164Q
splice site
g
V54L
S144F
Splice
R440X
EK470fs
Q519X
G629R
W784R
P795S
Q803X
S833F
LQ852X
Q912X
E962X
Q1010X
M1035K
L1095S
L1187F
Q1255X
L1411F
P1416L
S1443L
Splice
F1479L
Splice
K1691X
Q1822X
L2158X
E2193X
Q2228X
E2339X
Splice
2839 amino acids NF1 (ENST00000358273)
0
500
1000
1500
2000
2500
missense
nonsense
splice site
deletion
duplication
SV breaks

2

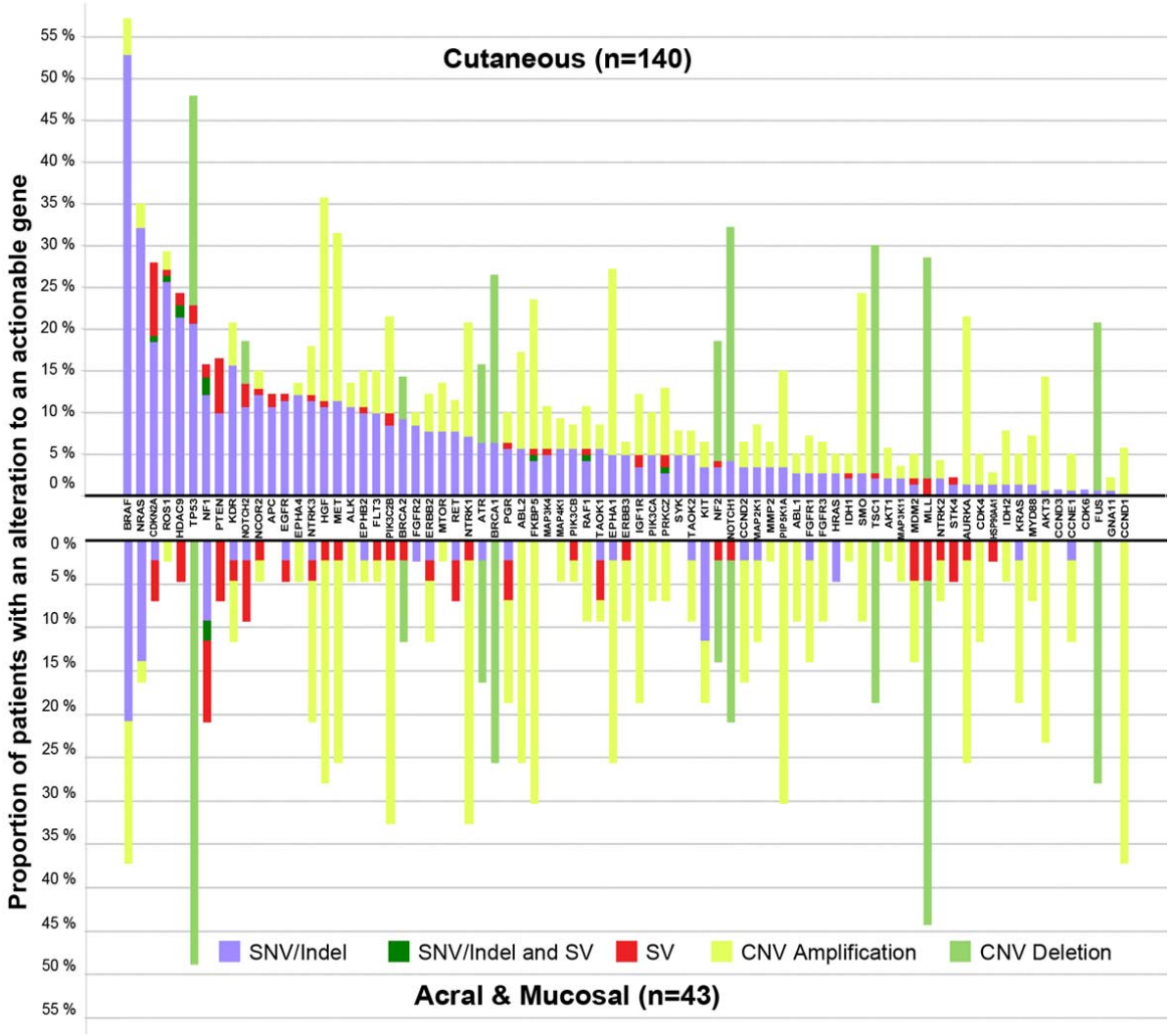
3

**Extended Data 8 | *BRAF*, *RAS* and *NF1* mutations. a and b,** *BRAF* somatic mutations were identified in 86 of the 183 samples (47%). V600E substitutions accounted for 48 (56%) of these mutations. Other activating mutations at codons 600 and 601 were less frequent (V600K, n=17; V600R, n=2*;* V600D, n=1; K601E, n=2). **c and d**, *NRAS* was mutated in 51 (28%) of tumours, 49 of these mutations occurred at 'hotspot' codons 12, 13 or 61. **e,** *HRAS* mutations occurred in 8 tumours, 5 of which were in 'hotspot' codons 12, 13 or 61. **f,** *KRAS* mutations occurred in 4 tumours, but none occurred at the commonly activating hotspot codons. **g,** *NF1* aberrations occurred throughout the gene in 32/183 (17%) tumours and included point mutations/small indels only (15 nonsense, 1 frameshift, 3 splice site and 4 missense). Structural rearrangements predicted to cause loss of function were frequent within *NF1* (structural rearrangement breakpoints indicated by back arrows). Some rearrangements were associated with a copy number change (green = loss; red = gain).




**Extended Data 9 | Candidate *RAF1* gene fusions.** Whole genome sequencing data was used to identify genomic rearrangements which could result in a gene fusion product. **a,** The RAF1 gene and protein are shown. **b,** A *RAF1* gene fusion was detected with a small transcript of *CDH3*. The gene fusion is expected to produce a protein which retains the ref1 kinase domain. **c,** A RAF1-

GOLGA4 gene fusion was detected which is expected to produce a protein which retains the ref1 kinase domain.




**Extended data 10 | Alterations to actionable genes in cutaneous and non-cutaneous melanomas.** Proportion of melanomas with an alteration to a gene that confers sensitivity to an FDA-approved or therapeutic agent being utilized in a clinical trial (in any cancer) in cutaneous and non-cutaneous melanomas. Patients were classified as having a nonsynonymous mutations/indels (SNV/indel, blue), both a non-silent mutation/indel and structural variant (SV, dark red), SV (light red) or copy number variation (CNV, high-level amplification orange and homozygous deletion green) to the specific gene. Genes are ordered by the difference between the frequency of mutated genes (excluding CNV) in cutaneous versus non-cutaneous melanomas, with genes with a higher mutation frequency in cutaneous melanomas on the left, moving to higher mutation frequencies in the non-cutaneous subtype on the right.